# Equal Contributions to Risk and Portfolio Construction

Master Thesis
by

David Stefanovits
stedavid@student.ethz.ch
ETH Zurich
8092 Zurich, Switzerland

Supervised by:
Paul Embrechts (ETH Zürich)
Frank Häusler (Wegelin Bank)
Urs Schubiger (Wegelin Bank)
Mario Wüthrich (ETH Zürich)

February 1 – June 30, 2010

Contents

Introduction

In this thesis we want to treat a non-traditional asset allocation approach, known in the financial world as *Equal Contributions to Risk* or *Risk Budgeting*. Portfolios constructed using this method are also referred to as *Risk Parity Portfolios*. This allocation principle relies on true *risk diversification* in order to limit the impact to the overall portfolios of losses in the individual instruments. The idea behind Equal Contributions to Risk is actually quite simple and intuitive. One golden rule in investment is: *don't put all your eggs in one basket*. Following this principle, an investor might choose to allocate his wealth in a portfolio of 50% stocks and 50% bonds. Balancing the investment based on the capital seems to agree with the "eggs" principle, however this portfolio does not offer true risk diversification. The problem is that stocks "eggs" are in fact much bigger than bonds "eggs" in terms of risk. This eggs analogy might appear simplistic, but it is not so far away from the real situation. For example, from 1925 until 2008, the S&P500 had an annualized volatility of 19.2% (see Section 4.2), whereas USA Bonds had an annualized volatility of 6.9%. Over this period, the observed sample correlation between the two is 0.1. The total portfolio volatility of a 50/50 portfolio is then given by (the following formulas will be justified in the first chapter)

$$\sigma_{50} = \sqrt{0.5^2 \cdot 0.192^2 + 0.5^2 \cdot 0.069^2 + 2 \cdot 0.5^2 \cdot 0.192 \cdot 0.069 \cdot 0.1} = 10.5\%,$$

and the contributions of the single assets to the total volatility of the portfolio are given by

$$\sigma_{\text{stocks}} = \frac{0.5 \cdot (0.5 \cdot 0.192^2 + 0.5 \cdot 0.192 \cdot 0.069 \cdot 0.1)}{\sigma_{50}} = 9.1\%$$
$$\sigma_{\text{bonds}} = \frac{0.5 \cdot (0.5 \cdot 0.069^2 + 0.5 \cdot 0.192 \cdot 0.069 \cdot 0.1)}{\sigma_{50}} = 1.4\%$$

Based on these computations, stocks contribute 86% of the volatility, and bonds contribute the remaining 14%. This shows that balancing the portfolio in terms of capital does not necessarily lead to a balanced portfolio from the perspective of volatility, i.e. risk. Risk Parity or Equal Contributions to Risk Portfolios are strategies where the exposures in the different financial instruments are chosen so that each instrument contributes equally to the total risk of the portfolios. In the previous example, we have considered the volatility as the portfolio risk, but this approach can theoretically be applied on a wide class of *risk measures*. In the first chapter, we introduce formally the concept of risk measures and risk contributions. How do we define meaningfully the risk contributions of single assets to the overall portfolio risk? Can we use the stand-alone risk of the individual assets? These questions, which are crucial in order to construct Risk Parity Portfolios, have been answered by Tasche in [24] and [25]. In Chapter 1, we formulate and explain these results that provide the mathematical foundations of the Equal Contributions to Risk Portfolio Construction. In the second chapter, we recall the classical portfolio optimization problems and formulate a new optimization problem, proposed by Maillard in [13], that corresponds to the Equal Contributions to Risk selection principle. We then explain in Chapter 3 how we are going to compute the risk contributions numerically. In particular, we focus on the problem of estimating the asset returns covariance matrix, which is very central in the context of this thesis. If we compare Equal Contributions to Risk with the classical mean-variance approach, we can observe that this selection principle is based only on risk diversification. Is this going to have a positive effect on the portfolio performance? We are going to discuss this issue in Chapter 4. We are going to construct Risk Parity strategies in some illustrative examples, and compare them with the classical strategies. Considering again the above example: the S&P500 annualized mean return over the period is 11.6%, and for the bonds is 5.9%. The Sharpe ratio of the stocks is therefore 0.6 and the one of the bonds is 0.86. The Sharpe ratio of the 50/50 strategy is hence 0.83, which is lower than that of bonds. This means that splitting our wealth equally in stocks and bonds is not increasing the return over risk performance than investing in bonds only. This indicates poor diversification. According to Qian [22], who is CIO of Panagora Asset Management (a Boston based Asset Management firm), the Sharpe ratio of Risk Parity Portfolios is higher than those of stocks and bonds, representing the benefits of diversification in terms of risk contributions. Hence, Qian claims that Equal Contributions to Risk Portfolios are not only efficient in terms of allocating risk but also in terms of performance. In Chapter 4, we would like to see if we can also observe such results in some examples using different risk measures and different distribution assumptions for the asset returns.

Chapter 1

Risk Attribution

1.1 Portfolio Model and Portfolio Risk

In this section we introduce the framework to describe mathematically the profit/loss generated by an investment consisting of several assets. We consider a *one period model* with fixed *time length* Δt for an investor who can invest in n financial *assets*. We use the term asset for stocks, bonds, derivatives, risky loans or similar financial instruments. Let $(\Omega, \mathcal{F}, P)$ be the probability space which represents the uncertainty about the future state of the market at the end of the period. Typically we will be interested in daily profits, i.e. $\Delta t = \frac{1}{365}$. This probability space is the domain of all the random variables we are going to introduce. For $i = 1, \ldots, n$ let the random variable R_i be the random *return* of asset i at the end of the time interval. We assume that the returns are random variable in $\mathcal{L}^1(\Omega, \mathcal{F}, P)$, i.e. $E[|R_i|] < \infty$ for $i = 1, \ldots, n$. Denote by $\mathbf{R} = (R_1, \ldots, R_n)$ the random *vector of returns*. As usual, positive returns correspond to profits and negative ones to losses. At the beginning of the time interval we take positions $\mathbf{m} = (m_1, \ldots, m_n) \in \mathbb{R}^n$ in the financial assets, where the *ith* element m_i denotes the amount of money invested in asset i. The vector $\mathbf{m}$ is called vector of *portfolio positions*, and a *portfolio* is represented by such a vector. The random *portfolio profit* at the the end of the time period, assuming that the portfolio positions remain unchanged during this period, is given by

$$X = \sum_{i=1}^{n} m_i R_i = \mathbf{m}'\mathbf{R}. \tag{1.1.1}$$

We also refer to X using the term *profit*, and we call the distribution of X *profit distribution*. Depending on the sign, this random variable indicates profit or loss caused by the investment. Usually we restrict ourselves only to some of all the possible positions: we define by $M \subset \mathbb{R}^n$ the set of portfolios under consideration. A typical choice for M is

$$M = \{\mathbf{m} = (m_1, \dots, m_n) \in \mathbb{R}^n_+ : \mathbf{m}'\mathbf{1} = W\}, \tag{1.1.2}$$

where we do not allow borrowing or short selling and we impose an initial wealth constraint of $W > 0$. We define by $\mathcal{M} = \{m'R : m \in M\} \subset \mathcal{L}^1(\Omega, \mathcal{F}, P)$ the set of all random variables which we interpret as portfolio profits. We now formalize mathematically the concept of portfolio risk.

Definition 1.1.1. *A risk measure on $\mathcal{M}$ is a mapping $\rho : \mathcal{M} \to \mathbb{R}$. We define ρ also on the set of portfolios $M \subset \mathbb{R}^n$ by setting $\rho(\mathbf{m}) := \rho(\mathbf{m}'\mathbf{R})$.*

We call $\rho(\mathbf{m})$ the *portfolio risk* and we interpret this quantity as the amount of capital that should be added to the portfolio $\mathbf{m}$ as a reserve in a risk-free asset in order to prevent insolvency. Positions with $\rho(\mathbf{m}) \leq 0$ are acceptable without injection of capital; if $\rho(\mathbf{m}) < 0$, we may even withdraw capital. Note that $\rho(\mathbf{m})$, as a reserve to compensate possible losses in the future, should be discounted with some factor depending on the risk-free interest rate. We do not care about this factor since we are only interested in losses relative to those of other portfolios. The above definition introduces the concept of risk measure as a general real-valued function on $\mathcal{M}$ (or M). Of course, if we want to represent risk realistically, we need to choose a function that satisfies special properties corresponding to our financial interpretation of risk. In fact, choosing an appropriate function ρ to represent the portfolio risk, is not an easy task, and some of the most popular choices used for decades by practitioners have been revealed inefficient in measuring risk. A "good" risk measure should at least satisfy the following properties. Note that many authors see for example [17] define positive values of X as future losses instead of profits of a position currently held. This leads to some sign differences in the properties.

Definition 1.1.2. *A risk measure $\rho : \mathcal{M} \to \mathbb{R}$ is called coherent on $\mathcal{M}$ if it satisfies the following properties.*

1. *For all $X \in \mathcal{M}, \lambda > 0$ with $\lambda X \in \mathcal{M}$: $\rho(\lambda X) = \lambda\rho(X)$ (positive homogeneity)*
2. *For all $X, Y \in \mathcal{M}$ with $X \leq Y$ a.s.: $\rho(X) \geq \rho(Y)$ (monotonicity)*
3. *For all $X \in \mathcal{M}, c \in \mathbb{R}$ with $X + c \in \mathcal{M}$: $\rho(X + c) = \rho(X) - c$ (translation invariance)*
4. *For all $X, Y \in \mathcal{M}$ with $X + Y \in \mathcal{M}$: $\rho(X + Y) \leq \rho(X) + \rho(Y)$ (subadditivity)*

The first property, which is a quite natural assumption, tells us that when all positions are increased by a multiple, the portfolio risk is also increased by the same multiple. This represents the fact that it is harder to liquidate larger positions and that the same portfolio does not allow for diversification. From the financial point of view the monotonicity is also a natural assumption: portfolios with almost surely lower profits than others must be more risky. The third property implies that investing in a risk-free asset reduces the amount of risk by exactly the value of the risk-free asset. This assumption is necessary in order to interpret $\rho(m)$ as a reserve capital: consider a portfolio m with profit X and $\rho(m) > 0$, adding the capital $\rho(m)$ to the position we get by translation invariance $\rho(X+\rho(X)) = \rho(X)-\rho(X) = 0$, and so the position $X + \rho(X)$ is acceptable. This is consistent with our interpretation of ρ. The subadditivity represents the diversification effect: "spreading out of investments should reduce risks". The use of a non-coherent risk measure is therefore arguable. We make some observations about consequences of the above properties. Note that the positive homogeneity implies that $\rho(0) = 0$, this follows because $\rho(0) = \rho(\lambda 0) = \lambda\rho(0)$ for all $\lambda > 0$. For a measure satisfying positive homogeneity, the monotonicity implies,

$$X \overset{(\geq)}{\leq} 0 \;\; a.s. \Rightarrow \rho(X) \overset{(\leq)}{\geq} 0. \tag{1.1.3}$$

This is obvious and intuitive: a zero position should have no risk, an almost surely negative profit should have positive risk and vice versa. Note that in the definition of risk measure and coherence, the set of portfolio profits $\mathcal{M}$ is part of the definition. We will sometimes put additional assumptions on the random variables in $\mathcal{M}$ when defining risk measures (see for example the standard deviation (1.2.1) where we consider only square integrable profits), and we will sometimes observe risk measures that are coherent only if we consider profits satisfying certain properties (see for example Theorem 1.2.1). Before focusing on the classical examples of risk measures used in practice, we want to consider a few typical distribution assumptions.

Example 1.1.1. A simple case is the assumption of normally distributed returns. As we are going to see, under this assumption the quantile based risk measures can be easily treated. However, it is well known that the assumption of normal distribution is questionable for stock market quotations: high losses are far more probable than the assumption of normal distribution. In the context of credit risk it is also clear that the asymmetric loss distribution by credit loans cannot be modelled by normal distribution. We assume that $\mathbf{R} \sim \mathcal{N}(\mu, \Omega)$, where $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\Omega \in \mathbb{R}^{n\times n}$, has a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and positive definite covariance matrix Ω (for a complete description of multivariate normally distributed random vectors see [17] 3.1.3). In this case, for a portfolio $\mathbf{m} \in M$, the profit X is also normally distributed with mean

$$E[X] = \sum_{i=1}^{n} m_i\mu_i = \mathbf{m}'\mu \tag{1.1.4}$$

and variance

$$Var[X] = \sum_{i=1}^{n}\sum_{j=1}^{n} m_i m_j Cov(R_i, R_j) = \mathbf{m}'\Omega\mathbf{m}. \tag{1.1.5}$$

The distribution function and density function of R_i are given by

$$F_{R_i} = \Phi\Big(\frac{x-\mu_i}{\sqrt{\Omega_{ii}}}\Big), \quad f_{R_i} = \frac{1}{\sqrt{\Omega_{ii}}}\varphi\Big(\frac{x-\mu_i}{\sqrt{\Omega_{ii}}}\Big) \tag{1.1.6}$$

and for X by

$$F_X = \Phi\Big(\frac{x-\mathbf{m}'\boldsymbol{\mu}}{\sqrt{\mathbf{m}'\Omega\mathbf{m}}}\Big), \quad f_X = \frac{1}{\sqrt{\mathbf{m}'\Omega\mathbf{m}}}\varphi\Big(\frac{x-\mathbf{m}'\boldsymbol{\mu}}{\sqrt{\mathbf{m}'\Omega\mathbf{m}}}\Big), \tag{1.1.7}$$

where $\Phi(x) = \int_{-\infty}^{x}\varphi(y)dy$ and $\varphi(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$ are the distribution and density functions of the standard normal distribution. The problem of estimating the model parameter Ω and $\boldsymbol{\mu}$ in practice will be considered in Chapter 3.

Example 1.1.2. In the Black-Scholes framework, we consider a portfolio of n stocks. The stock prices are modeled by a geometric Brownian motion with

$$S_i(t) = S_i(0)e^{\left(\mu_i - \frac{\sigma_i^2}{2}\right)t + \sigma_i W_i(t)} \tag{1.1.8}$$

for $t \geq 0$, $i = 1, \dots, n$ where the μ_i are the drift rates, σ_i are the volatilities and $\mathbf{W} = (W_1, \dots, W_n)$ being a multivariate standard Brownian motion with $Cov(W_i(t), W_j(t)) = \chi_{ij}t$ for $i \neq j$ and some correlation coefficient χ_{ij}. The returns are then given by

$$R_i = \frac{S_i(t+\Delta t) - S_i(t)}{S_i(t)} = e^{\left(\mu_i - \frac{\sigma_i^2}{2}\right)\Delta t + \sigma_i\left(W_i(t+\Delta t) - W_i(t)\right)} - 1, \tag{1.1.9}$$

and so we have

$$\log(R_i + 1) \sim \mathcal{N}\Big(\Big(\mu_i - \frac{\sigma_i^2}{2}\Big)\Delta t, \sigma_i^2\Delta t\Big), \tag{1.1.10}$$

i.e. R_i has a shifted lognormal distribution with mean $E[R_i] = e^{\mu_i\Delta t} - 1$, variance $Var[R_i] = (e^{\sigma_i^2\Delta t} - 1)e^{2\mu_i\Delta t}$, probability distribution function

$$F_{R_i}(r_i) = \Phi\left(\frac{\log(1+r_i) - \left(\mu_i - \frac{\sigma_i^2}{2}\right)\Delta t}{\sigma_i\sqrt{\Delta t}}\right) \tag{1.1.11}$$

and density function

$$f_{R_i}(r_i) = \frac{1}{\sigma_i \sqrt{\Delta t}(1+r_i)} \varphi\Big(\frac{\log(1+r_i) - \big(\mu_i - \frac{\sigma_i^2}{2}\big)\Delta t}{\sigma_i \sqrt{\Delta t}}\Big). \tag{1.1.12}$$

The distribution of $X = \mathbf{m}'R$, for a portfolio $\mathbf{m} \in M$, which is a linear combination of multivariate shifted lognormally distributed random variables, is not known explicitly. If we are interested in the distribution of X we need to use approximations. One possible way of doing this is to consider linear approximation of the function $\log(1+r)$ provided by the Taylor expansion, i.e. $\log(1+r) \approx r$ for r small. We are then in the multivariate normal case of the previous example:

$$R_i \sim \mathcal{N}\Big(\big(\mu_i - \frac{\sigma_i^2}{2}\big)\Delta t, \sigma_i^2 \Delta t\Big), \tag{1.1.13}$$

and

$$X \sim \mathcal{N}\Big(\Delta t \sum_{i=1}^n m_i\big(\mu_i - \frac{\sigma_i^2}{2}\big), m'\Omega m) \tag{1.1.14}$$

where $\Omega_{ij} := \sigma_i \sigma_j \chi_{ij} \Delta t$ for $i \neq j$ and $\Omega_{ii} := \sigma_i^2 \Delta t$.

Example 1.1.3. We consider a portfolio with n defaultable bonds without coupons. The returns are modeled by

$$R_i = r_i I_i - (1 - I_i) = (1 + r_i) I_i - 1, \tag{1.1.15}$$

where $r_i > 0$ is the return of the bond i if there is no default, and $I_i \in \{0, 1\}$ is the default indicator random variable with default probability $p_i = P[I_i = 0] \in (0, 1)$, for $i = 1, \ldots, n$. For simplicity we assume that the different bonds default independently. For the mean and the variance of the returns we have,

$$\begin{aligned} E[R_i] &= (1 + r_i)(1 - p_i) - 1, \\ Var[R_i] &= p_i(1 - p_i)(1 + r_i)^2. \end{aligned} \tag{1.1.16}$$

For the portfolio profit $X = m'R$, for a portfolio $m \in M$, we have

$$X = m'R \in \Big\{ \sum_{i \in U_1} m_i r_i - \sum_{i \in U_2} m_i \Big| U_1 \cup U_2 = \{1, \ldots, n\}, U_1 \cap U_2 = \emptyset \Big\}. \tag{1.1.17}$$

For a bipartition of the bonds U_1, U_2 with $U_1 \cup U_2 = \{1, \ldots, n\}$ and $U_1 \cap U_2 = \emptyset$, representing a possible future situation, we have using independence

$$
\begin{aligned}
P\Big[X = \sum_{i \in U_1} m_i r_i - \sum_{i \in U_2} m_i\Big] &= \prod_{i \in U_1} (1-p_i) \prod_{i \in U_2} p_i, \\
E[X] &= \sum_{i=1}^{n} m_i\big((1+r_i)(1-p_i) - 1\big), \qquad (1.1.18) \\
Var[X] &= \sum_{i=1}^{n} m_i^2 \Big(p_i(1-p_i)(1+r_i)^2\Big).
\end{aligned}
$$

In this kind of situations we typically have to deal with asymmetric loss distribtutions.

Example 1.1.4. It has been observed that the class of elliptical distributions provide better models for daily stock-return data than the multivariate normal distributions. The elliptical distributions, which generalize the multivariate normal distributions, are symmetric distributions that describe the returns much better especially in the tails. Examples of elliptical distributions include also t-distributions and logistic distributions. We present the basic theory of elliptical distributions in Appendix A. We illustrate this class of models considering the example of a multivariate t-distributed random vector: for $A \in \mathbb{R}^{n \times n}$ and $\boldsymbol{\mu} \in \mathbb{R}^n$,

$$
\mathbf{R} \overset{(d)}{\sim} A\mathbf{Y} + \boldsymbol{\mu}, \qquad (1.1.19)
$$

where $\mathbf{Y} = (Y_1, \ldots, Y_n)$ with $Y_i \overset{iid}{\sim} t(\nu)$, $\nu > 2$ for all $i = 1, \ldots, n$, i.e. the components of the random vector $\mathbf{Y}$ are independent t-distributed random variables with $\nu > 2$ degrees of freedom. The density and distribution function of the t-distribution with ν degrees of freedom are given by,

$$
\begin{aligned}
f_{Y_i}(y) = t_\nu(y) &= \frac{\Gamma\big(\frac{1}{2}(\nu+1)\big)}{\Gamma\big(\frac{1}{2}\nu\big)(\pi\nu)^{\frac{1}{2}}} \Big(1 + \frac{y^2}{\nu}\Big)^{-\frac{\nu+1}{2}}, \\
F_{Y_i}(y) = T_\nu(y) &= \frac{\Gamma\big(\frac{1}{2}(\nu+1)\big)}{\Gamma\big(\frac{1}{2}\nu\big)(\pi\nu)^{\frac{1}{2}}} \int_{-\infty}^{y} \Big(1 + \frac{t^2}{\nu}\Big)^{-\frac{\nu+1}{2}} dt,
\end{aligned} \qquad (1.1.20)
$$

where $\Gamma(\cdot)$ denotes the gamma function. Under the assumption $\nu > 2$ the first and second moments exist and we have

$$
\begin{aligned}
E[Y_i] &= 0, \\
Var[Y_i] &= \frac{\nu}{\nu-2}, \\
Cov(\mathbf{Y}) &= \frac{\nu}{\nu-2} I_n.
\end{aligned} \qquad (1.1.21)
$$

As we explain in Appendix A, the random vector $\mathbf{Y}$ is said to be spherically distributed if and only if its characteristic function $\phi_{\mathbf{Y}}(\mathbf{t})$ satisfies for all $\mathbf{t} \in \mathbb{R}^n$

$$\phi_{\mathbf{Y}}(\mathbf{t}) = \psi(\mathbf{t}'\mathbf{t}), \tag{1.1.22}$$

for a scalar function ψ. This is indeed the case for the random vector $\mathbf{Y}$, see [17] 3.7 and 3.21. Since this scalar function ψ describes completely the distribution of $\mathbf{Y}$, we use the notation $\mathbf{Y} \sim S_n(\psi)$ to say that $\mathbf{Y}$ is spherically distributed with *characteristic generator* ψ. Elliptical distributed random vectors are defined as affine transformations of spherically distributed random vectors. In this example $\mathbf{R}$ is elliptical distributed and for the characteristic function it holds by equation (A.4) that

$$\phi_R(\mathbf{t}) = e^{it'\boldsymbol{\mu}}\psi(\mathbf{t}'\Omega\mathbf{t}), \tag{1.1.23}$$

where $\Omega = A'A$. We use therefore the notation $\mathbf{R} \sim E_n(\boldsymbol{\mu}, \Omega, \psi)$ to indicate that $\mathbf{R}$ is elliptical distributed. Note that the matrix Ω is proportional to the covariance matrix of $\mathbf{R}$, as we can easily see,

$$Cov(\mathbf{R}) = Cov(A\mathbf{Y} + \mu) = Cov(A\mathbf{Y}) = ACov(\mathbf{Y})A' = \frac{\nu}{\nu - 2}\Omega. \tag{1.1.24}$$

If we assume that A is not singular we have,

$$\begin{aligned} f_R(r) &= \frac{1}{|det(A)|} f_Y\big(A^{-1}(r-\mu)\big) \\ &= \frac{\Gamma\big(\frac{1}{2}(\nu+n)\big)}{\Gamma\big(\frac{1}{2}\nu\big)(\pi\nu)^{\frac{n}{2}} det(\Omega)^{\frac{1}{2}}} \Big(1 + \frac{(r-\mu)'\Omega^{-1}(r-\mu)}{\nu}\Big)^{-\frac{\nu+n}{2}}. \end{aligned} \tag{1.1.25}$$

Since $X = \mathbf{m}'\mathbf{R}$, for $\mathbf{m} \in M$, is a linear combination of $\mathbf{R}$ we have that (see (A.5)) $\frac{X - m'\mu}{\sqrt{m'\Omega m}}$ is t-distributed and the density of X is given by

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{\mathbf{m}'\Omega\mathbf{m}}} t_\nu\bigg(\frac{1}{\sqrt{\mathbf{m}'\Omega\mathbf{m}}}\big(x - \mathbf{m}'\boldsymbol{\mu}\big)\bigg) \\ &= \frac{\Gamma\big(\frac{1}{2}(\nu+1)\big)}{\Gamma\big(\frac{1}{2}\nu\big)(\pi\nu)^{\frac{1}{2}}\sqrt{\mathbf{m}'\Omega\mathbf{m}}} \Big(1 + \frac{(x - \mathbf{m}'\boldsymbol{\mu})^2}{\nu\mathbf{m}'\Omega\mathbf{m}}\Big)^{-\frac{\nu+1}{2}}. \end{aligned} \tag{1.1.26}$$

This elliptical model provides a more robust parametric framework. High losses are described better by the heavier tails of the t-distribution than in the multivariate normal setting. The problem of estimating the model parameters will be considered in Chapter 3.

1.2 Examples of Risk Measures

In this section we present the major risk measures and discuss their properties.

1.2.1 Standard Deviation, Moments and Semi-Moments

The standard deviation is a very popular risk measure that was among others introduced by Markowitz in [14] to develop his portfolio theory. We assume that $Var[R_i] < \infty$ for $i = 1, \ldots, n$. We define the *standard deviation* risk measure of $X \in \mathcal{M}$ (or resp. $m \in M$) by

$$\sigma(X) = \sqrt{Var[X]} = \sqrt{Var[\mathbf{m}'\mathbf{R}]} = \sigma(\mathbf{m}). \tag{1.2.1}$$

Denoting by Ω the covariance matrix of $\mathbf{R}$ we have the simple representation

$$\sigma(\mathbf{m}) = \sqrt{\mathbf{m}'\Omega\mathbf{m}} = \sqrt{\sum_{i,j=1}^{n} m_i m_j \Omega_{ij}}. \tag{1.2.2}$$

A very popular variant of this measure is the so called *tracking error* with respect to a benchmark.

Example 1.2.1. The tracking error is defined as the standard deviation of the excess profit with respect to a benchmark. Denote by $\mathbf{b} = (b_1, \ldots, b_n) \in \mathbb{R}^n$ the portfolio positions of a benchmark. Let $\mathbf{e} = \mathbf{m} - \mathbf{b}$ be the vector of excess positions with respect to the benchmark. The tracking error is given by

$$TE(\mathbf{m}) = \sqrt{\mathbf{e}'\Omega\mathbf{e}}. \tag{1.2.3}$$

The standard deviation risk measure is always positive and so there is no position except for X a.s. constant that is acceptable without additional capital. It is then possible to fix a certain level $L > 0$ and define the following risk measure,

$$\rho_L(\mathbf{m}) = \sigma(\mathbf{m}) - L, \tag{1.2.4}$$

so that all positions with $\sigma(\mathbf{m}) \leq L$ are acceptable. This measure however is not positive homogeneous and therefore as we are going to see in the next section the Euler decomposition is not available. For this reason we are not going to use this measure. The standard deviation risk measure is quite simple to use analytically, but on the other hand it does not represent risk properly. Firstly, if we want to work with this measure we need that the second moments of the return exist, although this does not cause particular problems in finance. The main problem of using the standard deviation as risk measure is that both the fluctuation above and below mean are taken as contributions to risk. Since it makes no distinction between positive and negative deviations from the expected value, this is a good risk measure only for symmetric or roughly symmetric distributions, such as in the elliptical models. In areas, such as credit risk, where the distributions under consideration have high skewness, the use of standard deviation is problematic. The standard deviation is not a coherent risk measure in general. We discuss the properties. For a portfolio profit $X \in \mathcal{M}$ and $\lambda > 0$ such that $\lambda X \in \mathcal{M}$, we obviously have $\sigma(\lambda X) = \lambda\sigma(X)$, and so positive homogeneity is satisfied. The monotonicity and translation invariance clearly fail in general: the translation invariance fails because for any $c \in \mathbb{R}$ it holds $Var[X+c] = Var[X]$, the monotonicity because $X \leq Y$ a.s. does not imply any relation between σ_x and σ_y in general. The subadditivity is nevertheless satisfied: consider two square integrable profits $X, Y \in \mathcal{M}$ with $X+Y \in \mathcal{M}$, $Var[X] = \sigma_x^2$, $Var[X] = \sigma_y^2$ and $Cov(X,Y) = \sigma_{xy}$. The Cauchy-Schwarz Inequality for the covariance implies $\sigma_{xy} \leq \sigma_x\sigma_y$. Hence,

$$\sqrt{Var[X+Y]} = \sqrt{\sigma_x^2 + \sigma_y^2 + 2\sigma_{xy}} \leq \sqrt{\sigma_x^2 + \sigma_y^2 + 2\sigma_x\sigma_y} = \sigma_x + \sigma_y. \tag{1.2.5}$$

Since in this context we are interested in the lower part of the distribution of X (the part that corresponds to losses), a way of eliminating the deficiencies of the standard deviation risk measure, is to consider the lower semi-standard deviation of $X \in \mathcal{M}$ given by

$$\sigma^-(X) = \sigma^-(\mathbf{m}'\mathbf{R}) = \sqrt{E\left[\left(\left(\mathbf{m}'\mathbf{R} - E[\mathbf{m}'\mathbf{R}]\right)^-\right)^2\right]}, \tag{1.2.6}$$

where $\left(\mathbf{m}'\mathbf{R} - E[\mathbf{m}'\mathbf{R}]\right)^- = \max\{-\left(\mathbf{m}'\mathbf{R} - E[\mathbf{m}'\mathbf{R}]\right), 0\}$ denotes the negative part of the fluctuation. This measure however involves computational difficulties. It is also possible to consider higher moments as risk measures,

$$\begin{aligned} \sigma_k(X) &= \Big(E\big[(X - E[X])^k\big]\Big)^{\frac{1}{k}} \\ \sigma_k^-(X) &= \Big(E\big[((X - E[X])^-)^k\big]\Big)^{\frac{1}{k}}, \end{aligned} \tag{1.2.7}$$

for an integer k. For even central moments the higher we set k the more weight we give to deviations from the mean. Note that for a symmetric distributions around its mean the central moments with odd k are all equal to 0.

1.2.2 Value-At-Risk

Value-At-Risk, abbreviated VaR, has been accepted as risk measure in the last decade and has been frequently written into industrial regulations. The

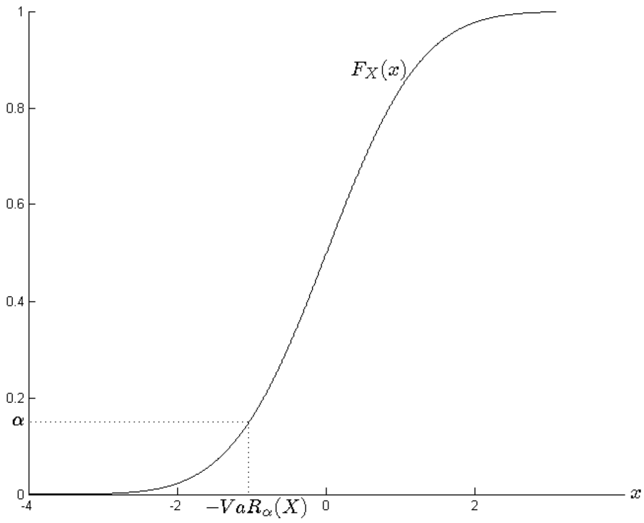


Figure 1.1: Value-At-Risk for a nice distribution.

idea behind this way of measuring risk is easy to understand, but presents some theoretical deficiencies.

Definition 1.2.1. *For $\alpha \in (0,1)$ and $X \in \mathcal{M}$ we define the lower α-quantile of X by*

$$q_\alpha(X) = \inf\{x \in \mathbb{R} : P[X \leq x] \geq \alpha\}. \tag{1.2.8}$$

The VaR is defined as the negative of the lower α-quantile of the portfolio profit distribution,

$$VaR_\alpha(X) = -q_\alpha(X). \tag{1.2.9}$$

The VaR can be interpreted as the amount of capital needed as reserve in order to prevent insolvency which happens with probability α. We refer to α as the *confidence level*, and typical values set for this probability are $\alpha = 99\%$ or $\alpha = 95\%$. Note that if the distribution function F_X is continuous and strictly increasing, then it holds

$$VaR_\alpha(X) = -F_X^{-1}(\alpha), \tag{1.2.10}$$

where F_X^{-1} denotes the inverse function of the distribution function of X. A simple equivalent criterion to find the lower α-quantile is the following.

Lemma 1.2.1. *A point $q \in \mathbb{R}$ is the lower α-quantile of some distribution function F if and only if $F(q) \geq \alpha$ and $F(x) \leq \alpha$ for all $x < q$.*

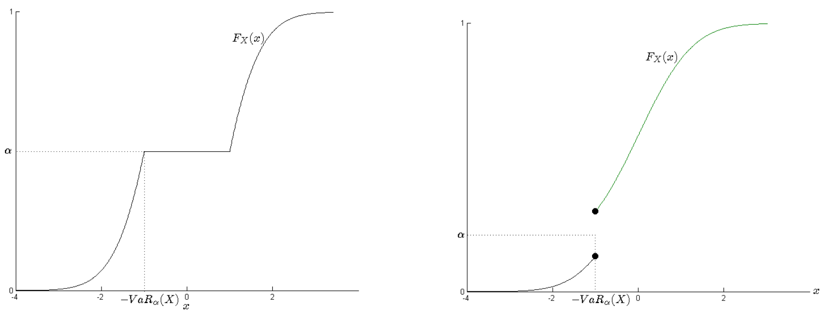


Figure 1.2: Value-At-Risk in two tricky situations.

Example 1.2.2. Consider the simple case of multivariate normal distributed returns described in Example 1.1.1. Since in this case the portfolio profit X has a normal distribution with mean $m'\mu$ and variance $m'\Omega m$, the lower α-quantile is given by the following equation (1.21),

$$\Phi\Big(\frac{q_\alpha(\mathbf{m'R}) - \mathbf{m'}\boldsymbol{\mu}}{\sqrt{\mathbf{m'}\Omega\mathbf{m}}}\Big) = \alpha \tag{1.2.11}$$

and so $VaR_\alpha(X) = -\sqrt{\mathbf{m'}\Omega\mathbf{m}}\Phi^{-1}(\alpha) - \mathbf{m'}\boldsymbol{\mu}$.
In the same way we can compute VaR in the elliptical case of Example 1.1.4, where we have $\frac{X-\mathbf{m'}\boldsymbol{\mu}}{\sqrt{\mathbf{m'}\Omega\mathbf{m}}} \sim t(\nu)$, and so

$$VaR_\alpha(X) = -\sqrt{\mathbf{m'}\Omega\mathbf{m}}T_\nu^{-1}(\alpha) - \mathbf{m'}\boldsymbol{\mu}. \tag{1.2.12}$$

There are some technical difficulties when using quantile based risk measures, such as VaR, which is worth to explain (see Figure 1.2). Consider for example a probability distribution function F_X that has a jump by x_0 with

$$\lim_{x\uparrow x_0} F_X(x) = \alpha_- < \alpha_+ = \lim_{x\downarrow x_0} F_X(x) = F_X(x_0) \tag{1.2.13}$$

In this case we have many confidence levels giving the same Value-At-Risk:

$$\forall \alpha \in (\alpha_-, \alpha_+) : VaR_\alpha(X) = x_0. \tag{1.2.14}$$

Consider also the case of a distribution function that is constant on the interval $[x_-, x_+)$ with probability α. In this case we have $VaR_\alpha(X) = x_-$. These examples correspond to the two critical cases where the equation $P(X \le x) = \alpha$ has no solution or a whole range of solutions.
VaR is not coherent in general. We examine the four properties. The homogeneity and translation invariance are clearly satisfied for this risk measure. The monotonicity is also clearly satisfied: for $X \le Y$ almost surely we have for any $\alpha \in (0, 1)$

$$P[Y \le y] \ge \alpha \Rightarrow P[X \le y] \ge \alpha \tag{1.2.15}$$

and so taking the infimum it follows $q_\alpha(Y) \ge q_\alpha(X)$, which implies the desired inequality. The subadditivity property fails in general as shown by the following simple example.

Example 1.2.3. Consider two independent payoffs X and Y with $X, Y \in [-4, -2] \cup [1, 2]$ and

$$P(X < 0) = P(Y < 0) = 3\%. \tag{1.2.16}$$

For the combined position with payoff $X + Y$ we have (because of the image of X, Y and independence)

$$\begin{aligned} P(X + Y < 0) = P(X < 0, Y > 0) + P(X > 0, Y < 0) + \\ + P(X < 0, Y < 0) = 5.91\% \end{aligned} \tag{1.2.17}$$

and this implies $q_{5\%}(X + Y) < 0$, i.e. $VaR_{5\%}(X + Y) > 0$. We also have $VaR_{5\%}(X), VaR_{5\%}(Y) \le 0$ and so

$$VaR_{5\%}(X + Y) > VaR_{5\%}(X) + VaR_{5\%}(Y), \tag{1.2.18}$$

subadditivity is thus not available.

In the literature we can find other examples of more realistic situations than the above where we do not have subadditivity. See for example [17] 6.7, where the authors show the lacking of subadditivity of VaR for a portfolio of defaultable bonds. Also in [8] there is a less sophisticated counterexample. Nevertheless, in the situation of elliptical distributed asset returns we have the following result ([17] Theorem 6.8).

Theorem 1.2.1. *Let $R \sim E_n(\mu, \Omega, \psi)$ and $0 < \alpha \le \frac{1}{2}$. Then VaR is subadditive on $\mathcal{M}$ and thus coherent.*

Proof. Let $X_1, X_2 \in \mathcal{M}$ be two portfolio profits such that $X_1 + X_2 \in \mathcal{M}$. Using the definition of elliptical distributed random vectors (see Appendix A Definition A.2) we can write

$$X_1 = \mathbf{m_1}'\mathbf{R} = \mathbf{m_1}'A\mathbf{Y} + \mathbf{m_1}'\boldsymbol{\mu}, \tag{1.2.19}$$

for $\mathbf{Y} \sim S_k(\psi)$, $A \in \mathbb{R}^{n \times k}$. By Theorem A.1

$$X_1 \overset{(d)}{\sim} |\mathbf{m_1}'A| Y_1 + \mathbf{m_1}'\boldsymbol{\mu}, \tag{1.2.20}$$

where $|\mathbf{m_1}'A|$ denotes the euclidean norm of the vector $\mathbf{m_1}'A$ and Y_1 the first component of $\mathbf{Y}$. The translation invariance and homogeneity of VaR imply

$$VaR_\alpha(X_1) = |\mathbf{m_1}'A|VaR_\alpha(Y_1) + \mathbf{m_1}'\boldsymbol{\mu}. \tag{1.2.21}$$

In the same way we have $VaR_\alpha(X_2) = VaR_\alpha(\mathbf{m_2}'\mathbf{R}) = |\mathbf{m_2}'A|VaR_\alpha(Y_1) + \mathbf{m_2}'\boldsymbol{\mu}$ and $VaR_\alpha(X_1+X_2) = |(\mathbf{m_1}+\mathbf{m_2})'A|VaR_\alpha(Y_1)+(\mathbf{m_1}+\mathbf{m_2})'\boldsymbol{\mu}$. Since Y_1 is spherically distributed and must be symmetric we have $VaR_\alpha(Y_1) \geq 0$ for $\alpha \leq \frac{1}{2}$,

$$\begin{aligned} VaR_\alpha(X_1+X_2) &= \underbrace{|(\mathbf{m_1}+\mathbf{m_2})'A|}_{\leq|\mathbf{m_1}'A|+|\mathbf{m_2}'A|}\underbrace{VaR_\alpha(Y_1)}_{\geq 0}+(\mathbf{m_1}+\mathbf{m_2})'\boldsymbol{\mu} \\ &\leq VaR_\alpha(X_1) + VaR_\alpha(X_2). \end{aligned} \tag{1.2.22}$$

□

See also [7] for more interpretation on this result. Since VaR in general does not have this diversification effect, this measure may lead to unrealistic results. An investor could for example be encouraged to split his account into two in order to meet the lower capital requirement. Another big issue about this way of measuring the portfolio risk is that VaR provides a minimum bound for losses ignoring potential large losses beyond this limit.

1.2.3 Expected Shortfall

The following three risk measures define alternatives to VaR that take the tail losses into account.

Definition 1.2.2. *Assume that $E[X^-] < \infty$ for all $X \in \mathcal{M}$. For $\alpha \in (0,1)$ we define the following risk measures on $\mathcal{M}$*

1. *The lower α-tail conditional expectation is defined as*

$$TCE_\alpha(X) = -E[X|X \leq -VaR_\alpha(X)]. \tag{1.2.23}$$

2. *We define the lower α-tail distribution of the profit by the following distribution function*

$$F_X^\alpha(x) = \begin{cases} 0, & x < VaR_\alpha(X) \\ \frac{\alpha - F(-x)}{\alpha}, & x \geq VaR_\alpha(X), \end{cases} \tag{1.2.24}$$

   we define the conditional Value-at-Risk, $CVaR_\alpha(X)$, as as the mean of the lower α-tail distribution of X.

3. *We define the expected shortfall at level α by*

$$\begin{aligned} ES_\alpha(X) = -\frac{1}{\alpha}\Big\{ & E\big[X 1_{\{X \le -VaR_\alpha(X)\}}\big] - \\ & - VaR_\alpha(X)\big(\alpha - P[X \le -VaR_\alpha(X)]\big)\Big\} = \\ & = -\frac{1}{\alpha}\Big\{E\big[X 1_{\{X \le q_\alpha(X)\}}\big] + q_\alpha(X)\big(\alpha - P[X \le q_\alpha(X)]\big)\Big\}. \end{aligned} \tag{1.2.25}$$

It is evident from the definition that these three risk measures are in a very close relation. TCE was proposed by Artzner et al. in [2] and it was shown by the authors that this measure is in general not subadditive, thus not coherent. However, in case the profit distribution has no jump at $q_\alpha(X)$, the following result follows from the above definitions.

Theorem 1.2.2. *Assume that for all $X \in \mathcal{M}$ we have $P[X \le q_\alpha(X)] = \alpha$. Then, for any $X \in \mathcal{M}$*

$$\begin{aligned} ES_\alpha(X) &= -\frac{1}{P[X \le -VaR_\alpha(X)]} E\big[X 1_{X \le -VaR_\alpha(X)}\big] = \\ &= -E[X | X \le -VaR_\alpha(X)] = TCE_\alpha(X), \end{aligned} \tag{1.2.26}$$

i.e. the tail conditional expectation and the expected shortfall are the same, and TCE_α defines a coherent risk measure on $\mathcal{M}$.

For the proof we need the following elementary property.

Lemma 1.2.2. *Let $Y \in \mathcal{L}^1(\Omega, \mathcal{F}, P)$ and $y \in \mathbb{R}$ such that $P[Y \le y] > 0$. For any event $F \in \mathcal{F}$ such that $P[F] \ge P[Y \le y]$ it holds*

$$E[Y|F] \ge E[Y|Y \le y] \tag{1.2.27}$$

Proof. Assume first that $P\big[F \cap \{Y \le y\}\big] = 0$. Then, $Y > y$ P-a.s. on F, and so $E[Y|F] \ge y \ge E[Y|Y \le y]$. Assume now that $P\big[F \cap \{Y \le y\}\big] > 0$. We have,

$$
\begin{aligned}
E[Y|Y\le y] &= y + \frac{1}{P[Y\le y]}\Big(E\big[(Y-y)1_{\{Y\le y\}\cap F}\big] + \underbrace{E\big[(Y-y)1_{\{Y\le y\}\cap(\Omega\setminus F)}\big]}_{\le 0}\Big)\\
&\le y + \underbrace{E\big[Y-y\big|\{Y\le y\}\cap F\big]}_{\le 0}\,\underbrace{P[F|Y\le y]}_{\ge P[Y\le y|F]}\\
&\le y + E\big[Y-y\big|\{Y\le y\}\cap F\big]P[Y\le y|F]\\
&= y + \frac{E\big[(Y-y)1_{\{Y\le y\}\cap F}\big]}{P[F]}\\
&\le y + \frac{E\big[(Y-y)1_{\{Y\le y\}\cap F}\big]}{P[F]} + \frac{E\big[(Y-y)1_{\{Y> y\}\cap F}\big]}{P[F]}\\
&= E[Y|F],
\end{aligned}
\tag{1.2.28}
$$

which is the desired inequality. □

Proof. We now prove Theorem 1.2.2. Since for any $X \in \mathcal{M}$ we have $P[X \le q_\alpha(X)] = \alpha$ we can immediately see that $ES_\alpha(X) = TCE_\alpha(X)$. For the coherence: positive homogeneity and translation invariance are clear. For the monotonicity consider $X \le Y$ a.s.. By assumption we have $P[X \le q_\alpha(X)] = P[Y \le q_\alpha(Y)] = \alpha$ and applying Lemma 1.2.2 we have,

$$
\begin{aligned}
\rho(X) = -E[X|X\le q_\alpha(X)] &\ge -E[X|Y\le q_\alpha(Y)]\\
&\ge -E[Y|Y\le q_\alpha(Y)] = \rho(Y).
\end{aligned}
\tag{1.2.29}
$$

For the subadditivity applying again Lemma 1.2.2:

$$
\begin{aligned}
\rho(X+Y) &= -E[X|X+Y\le q_\alpha(X+Y)] - E[Y|X+Y\le q_\alpha(X+Y)]\\
&\le -E[X|X\le q_\alpha(X)] - E[Y|Y\le q_\alpha(Y)] = \rho(X)+\rho(Y).
\end{aligned}
\tag{1.2.30}
$$

□

We defined CVaR, as in the paper [23] of Rockafellar and Uryasev, based on a rescaled probability distribution that focus only on the lower tail part of the original distribution. From the definition above it is evident that F_X^α defines a probability distribution. We represent graphically how the lower α-tail distribution is constructed.
ES generalizes TCE for the case of a jump at q_α. As we are going to see, this modification of TCE defines a coherent risk measure. In [23] the authors discuss these definitions in details showing that CVaR and ES are actually the same. In this section we want to discuss the properties of ES

(or CVaR). A first very useful result is the following integral representation of the expected shortfall ([1] Proposition 3.2).

Theorem 1.2.3. *Assume that* $E[X^-] < \infty$. *Then,*

$$ES_\alpha(X) = \frac{1}{\alpha}\int_0^\alpha VaR_u(X)du. \tag{1.2.31}$$

Proof. Consider a uniformly distributed random variable on $[0,1]$ and define the random variable $Z = q_U(X)$. Note that Z has the same distribution as X, since

$$P[Z \le x] = P[q_U(X) \le x] = P[F_X^{-1}(U) \le x] = P[U \le F_X(x)] = F_X(x). \tag{1.2.32}$$

The map $u \mapsto q_u(X)$ is obviously non-decreseaing and therefore we have the following sets relations

$$\begin{aligned} &\{U \le \alpha\} \subset \{q_U(X) \le q_\alpha(X)\}, \\ &\{U > \alpha\} \cap \{q_U(X) \le q_\alpha(X)\} \subset \{q_U(X) = q_\alpha(X)\}, \end{aligned} \tag{1.2.33}$$

which implies

$$\{U \le \alpha\} \cup \big(\{U > \alpha\} \cap \{q_U(X) \le q_\alpha(X)\}\big) = \{q_U(X) \le q_\alpha(X)\}, \tag{1.2.34}$$

where the union is disjoint. Then,

$$\begin{aligned} \int_0^\alpha q_u(X)du &= E\big[q_U(X)1_{\{U\le\alpha\}}\big] \\ &\overset{(1.2.32)}{=} E\big[q_U(X)1_{\{q_U(X)\le q_\alpha(X)\}}\big] - E\big[q_U(X)1_{\{U>\alpha\}\cap\{q_U(X)\le q_\alpha(X)\}}\big] \\ &\overset{(1.2.30)}{=} E\big[X1_{\{X\le q_\alpha(X)\}}\big] - E\big[X1_{\{U>\alpha\}\cap\{X\le q_\alpha(X)\}}\big] \\ &\overset{(1.2.31)}{=} E\big[X1_{\{X\le q_\alpha(X)\}}\big] - q_\alpha(X)P\big[\{U>\alpha\}\cap\{X\le q_\alpha(X)\}\big] \\ &\overset{(1.2.32)}{=} E\big[X1_{\{X\le q_\alpha(X)\}}\big] - q_\alpha(X)\big(\alpha - P[\{X\le q_\alpha(X)\}]\big). \end{aligned} \tag{1.2.35}$$

Dividing both sides by $-\alpha$ we get the desired identity. □

We now prove that the expected shortfall is a proper risk measure.

Theorem 1.2.4. *Assume* $E[X^-] < \infty$ *for all* $X \in \mathcal{M}$. ES_α *defines a coherent risk measure on* $\mathcal{M}$.

Proof. 1. For $\lambda > 0$ we have

$$ES_\alpha(\lambda X) = -\frac{1}{\alpha}\Big\{\lambda E\big[X1_{\{X\le q_\alpha(X)\}}\big] + \lambda q_\alpha(X)\big(\alpha - P[\lambda X \le \lambda q_\alpha(X)]\big)\Big\}$$
$$= \lambda ES_\alpha(X) \tag{1.2.36}$$

2. The monotonicity follows immediately by the monotonicity of VaR and the integral representation in Theorem 1.2.3.

3. Using the fact that $q_\alpha(X+c) = q_\alpha(X) + c$ we have

$$\begin{aligned} ES_\alpha(X+c) = -\frac{1}{\alpha}\Big\{&E[X1_{X\le q_\alpha(X)}] + cP[X \le q_\alpha(X)] \\ &+ (q_\alpha(X)+c)\big(\alpha - P[X\le q_\alpha(X)]\big)\Big\} \\ = ES_\alpha&(X) - c \end{aligned} \tag{1.2.37}$$

4. Define the following random variable

$$1^\alpha_{\{X\le x\}} = \begin{cases} 1_{\{X\le x\}} & \text{if } \ P[X=x] = 0, \\ 1_{\{X\le x\}} + \frac{\alpha - P[X\le x]}{P[X=x]}1_{\{X=x\}} & \text{if } \ P[X=x] > 0. \end{cases} \tag{1.2.38}$$

This random variable satisfies

$$E[1^\alpha_{\{X\le q_\alpha(X)\}}] = \alpha. \tag{1.2.39}$$

If $P[X = q_\alpha(X)] > 0$ then integrating (1.2.34) gives directly α. If $P[X = q_\alpha(X)] = 0$ we have $E[1^\alpha_{\{X\le q_\alpha(X)\}}] = P[X \le q_\alpha(X)] \ge \alpha$, on the other hand in this case we have $P[X \le q_\alpha(X)] = P[X < q_\alpha(X)] \le \alpha$ (see Lemma 1.2.1) and therefore the expectation must be equal to α. It also holds

$$E[1_{\{X\le q_\alpha(X)\}^\alpha}] \in [0,1], \tag{1.2.40}$$

this follows because $P[X \le q_\alpha(X)] \ge \alpha$. We have also by definition

$$\frac{1}{\alpha}E[X1^\alpha_{\{X\le q_\alpha(X)\}}] = -ES_\alpha(X). \tag{1.2.41}$$

We also have by (1.2.36)

$$\begin{cases} 1^\alpha_{\{Z\le q_\alpha(Z)\}} - 1^\alpha_{X\le q_\alpha(X)} \ge 0 & \text{if } \ X > q_\alpha(X), \\ 1^\alpha_{\{Z\le q_\alpha(Z)\}} - 1^\alpha_{X\le q_\alpha(X)} \le 0 & \text{if } \ X < q_\alpha(X), \end{cases} \tag{1.2.42}$$

where we set $Z := X + Y$. Using this we can estimate

$$\begin{aligned}
&ES_\alpha(X) + ES_\alpha(Y) - ES_\alpha(Z) = \\
&= \frac{1}{\alpha}E\big[Z1^\alpha_{\{Z\le q_\alpha(Z)\}} - X1^\alpha_{\{X\le q_\alpha(X)\}} - Y1^\alpha_{\{Y\le q_\alpha(Y)\}}\big] \\
&= \frac{1}{\alpha}E\Big[X\big(1^\alpha_{\{Z\le q_\alpha(Z)\}} - 1^\alpha_{\{X\le q_\alpha(X)\}}\big) + Y\big(1^\alpha_{\{Z\le q_\alpha(Z)\}} - 1^\alpha_{\{Y\le q_\alpha(Y)\}}\big)\Big] \\
&\ge \frac{1}{\alpha}\Big(q_\alpha(X)\underbrace{E\big[1^\alpha_{\{Z\le q_\alpha(Z)\}} - 1^\alpha_{\{X\le q_\alpha(X)\}}\big]}_{=0} \\
&+ q_\alpha(Y)\underbrace{E\big[1^\alpha_{\{Z\le q_\alpha(Z)\}} - 1^\alpha_{\{Y\le q_\alpha(Y)\}}\big]}_{=0}\Big) \\
&= 0,
\end{aligned} \tag{1.2.43}$$

which is subadditivity.

□

We now look at formulas for the ES in the case of the two main distribution assumptions we have considered so far.

Example 1.2.4. We can compute the expected shortfall for the two main distribution situations we have examined so far, in the case of the multivariate normal model we have $\frac{X-\mathbf{m}'\boldsymbol{\mu}}{\sqrt{\mathbf{m}'\Omega\mathbf{m}}} \sim \mathcal{N}(0,1)$ and by linearity of expectation

$$\begin{aligned}
ES_\alpha(X) &= -E[X|X \le -VaR_\alpha(X)] = \\
&= -\mathbf{m}'\boldsymbol{\mu} - \sqrt{\mathbf{m}'\Omega\mathbf{m}}E\left[\frac{X-\mathbf{m}'\boldsymbol{\mu}}{\sqrt{\mathbf{m}'\Omega\mathbf{m}}}\middle|\frac{X-\mathbf{m}'\boldsymbol{\mu}}{\sqrt{\mathbf{m}'\Omega\mathbf{m}}} \le -VaR_\alpha\left(\frac{X-\mathbf{m}'\boldsymbol{\mu}}{\sqrt{\mathbf{m}'\Omega\mathbf{m}}}\right)\right] \\
&= -\mathbf{m}'\boldsymbol{\mu} - \frac{\sqrt{\mathbf{m}'\Omega\mathbf{m}}}{\alpha}\int_{-\infty}^{\Phi^{-1}(\alpha)} x\varphi(x)dx \\
&= -\mathbf{m}'\boldsymbol{\mu} - \frac{\sqrt{\mathbf{m}'\Omega\mathbf{m}}}{\alpha}\big[-\varphi(x)\big]_{-\infty}^{\Phi^{-1}(\alpha)} \\
&= -\mathbf{m}'\boldsymbol{\mu} + \frac{\sqrt{\mathbf{m}'\Omega\mathbf{m}}}{\alpha}\varphi\big(\Phi^{-1}(\alpha)\big).
\end{aligned} \tag{1.2.44}$$

Similarly for the case where $\frac{X-\mathbf{m}'\boldsymbol{\mu}}{\sqrt{\mathbf{m}'\Omega\mathbf{m}}} \sim t(\nu)$ we have

$$
\begin{aligned}
ES_\alpha(X) &= -\mathbf{m}'\boldsymbol{\mu} - \frac{\sqrt{\mathbf{m}'\Omega\mathbf{m}}}{\alpha}\int_{-\infty}^{T_\nu^{-1}(\alpha)} x t_\nu(x)dx \\
&= -\mathbf{m}'\boldsymbol{\mu} - \frac{\sqrt{\mathbf{m}'\Omega\mathbf{m}}}{\alpha}\left[\frac{\nu}{1-\nu}\Big(1+\frac{x^2}{\nu}\Big)t_\nu(x)\right]_{-\infty}^{T_\nu^{-1}(\alpha)} \\
&= -\mathbf{m}'\boldsymbol{\mu} - \frac{\nu}{1-\nu}\frac{\sqrt{\mathbf{m}'\Omega\mathbf{m}}}{\alpha}\left(1+\frac{\big(T_\nu^{-1}(\alpha)\big)^2}{\nu}\right)t_\nu\big(T_\nu^{-1}(\alpha)\big).
\end{aligned}
\tag{1.2.45}
$$

Comparing the ratio $ES_\alpha(X)/Var_\alpha(X)$ in the two cases, it is possible to show that in the multivariate normal model we have that

$$\frac{ES_\alpha(X)}{Var_\alpha(X)} \longrightarrow 1 \tag{1.2.46}$$

as $\alpha \to 0$, whereas in the elliptical model we have

$$\frac{ES_\alpha(X)}{Var_\alpha(X)} \longrightarrow \nu/(\nu-1) > 1 \tag{1.2.47}$$

as $\alpha \to 0$ (see [16] Section 4.2). The difference between VaR and ES is greater in the elliptical case because of the heavy tails of the loss distribution.

The following characterization proved by Acerbi and Tasche in [1] allows us to think about ES as the limiting average of the worst $100\alpha\%$ losses. This result can be used in practice to construct Monte Carlo estimates for the risk.

Lemma 1.2.3. *For $\alpha \in (0,1)$, $X \in \mathcal{M}$ with $E[X^-] < \infty$, and $(X_i)_{i\in\mathbb{N}}$ a sequence of independent random variable distributed as X, we have*

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{\lfloor n\alpha \rfloor} X_{i:n}}{\lfloor n\alpha \rfloor} = ES_\alpha(X) \tag{1.2.48}$$

almost surely, where $\lfloor n\alpha \rfloor$ denotes the biggest integer smaller than $n\alpha$ and $X_{1:n} \leq \cdots \leq X_{n:n}$ denote the components of the ordered n-tuple $(X_1, \ldots, X_n)$.

1.3 Euler's Relation and Risk Contributions

The main problem we want to focus on now is to determine the risk of individual securities in order to construct trading strategies that take the single risks of the instruments into account. It is however a quite hard task to identify the risk of each source in large and complex portfolios. This is because the risk of individual securities measured by most common risk measures such as the ones we have discussed above do not sum up to the total risk of the portfolio. The stand-alone risk of an individual asset is very significant, but it contributes little to the overall risk of the portfolio because of the correlations with other assets. Therefore, the stand-alone risks of the assets do not provide much information about the composition of the total portfolio risk, and so cannot be used to construct portfolios. This is closely related to the concept of diversification, see Definition 1.1.2 of a coherent risk measure.

Example 1.3.1. We illustrate this in a very basic bivariate normal situation. Consider a portfolio consisting two correlated assets with returns $R_i \sim \mathcal{N}(\mu_i, \sigma_i)$ for $i = 1, 2$ and $Cov(R_1, R_2) = \chi$. The individual VaR's per unit are $VaR_\alpha(R_i) = -\sigma_i \Phi^{-1}(\alpha) - \mu_i$. If we assume that $\chi < \sigma_1 \sigma_2$ and that $\alpha \in (0, \frac{1}{2})$ we have for the risk of the portfolio with one unit of each asset

$$VaR_\alpha(R_1 + R_2) < VaR_\alpha(R_1) + VaR_\alpha(R_2). \tag{1.3.1}$$

which shows that individual risks cannot fully describe the portfolio risk.

We want to introduce a meaningful methodology to decompose the total risk of a portfolio into *risk contributions* of the individual financial instruments in the portfolio. The methodology applies to a wide class of risk measures.

Definition 1.3.1. *A risk measure $\rho : \mathcal{M} \to \mathbb{R}$ is said to be positive homogeneous of degree τ if for all $X \in \mathcal{M}, \lambda > 0$ with $\lambda X \in \mathcal{M}$ we have $\rho(\lambda X) = \lambda^\tau \rho(X)$. If ρ satisfies this with $\tau = 1$ we say as in Definition 1.1.2 that ρ is positive homogeneous.*

The standard deviation, VaR, ES and all coherent risk measures belong to this class with $\tau = 1$. The following result, which is actually based only on elementary calculus, provides the desired decomposition. Note that, since a risk measure can be viewed as a function on the subset $M \subset \mathbb{R}^n$ of portfolios, the notion of differentiability is meaningful when referred to a risk measure.

Theorem 1.3.1. *Let ρ be a positive homogeneous risk measure on M of degree τ, and assume that M is an open set such that for all $m \in M, \lambda > 0$ we have $\lambda m \in M$. If ρ is partially differentiable on M with respect to the m_i's , then for all $m \in M$*

$$\rho(\mathbf{m}) = \frac{1}{\tau} \sum_{i=1}^{n} m_i \frac{\partial \rho}{\partial m_i}(\mathbf{m}). \tag{1.3.2}$$

Proof. Consider the real map

$$\mathbb{R}_+ \ni t \mapsto \rho(t\mathbf{m}) = \rho(tm_1, \ldots, tm_n) \in \mathbb{R}. \tag{1.3.3}$$

Applying the chain rule for differentiable functions in several variables we get

$$\frac{d\rho(\lambda\mathbf{m})}{d\lambda} = \sum_{i=1}^{n} \frac{\partial\rho(\lambda\mathbf{m})}{\partial m_i} m_i. \tag{1.3.4}$$

Because of homogeneity the left hand side becomes

$$\frac{d\rho(\lambda\mathbf{m})}{d\lambda} = \frac{d}{d\lambda}\big(\lambda^\tau \rho(\mathbf{m})\big) = \tau\lambda^{\tau-1}\rho(\mathbf{m}), \tag{1.3.5}$$

and setting $\lambda = 1$ leads to (1.3.2). □

The above decomposition is fundamental when considering risk contributions and can be applied to all the risk measures seen in Section 1.2, since all of them are positive homogeneous measures. By quantile based risk measures we need however some more assumptions on the distribution of the assets returns in order to ensure differentiability. We will discuss this aspect in details in Section 1.4. Each component $m_i \frac{\partial\rho(\mathbf{m})}{\partial m_i}$, termed *risk contribution of asset* i, is the amount of risk contributed to the total risk by investing m_i in asset i. In the case $\tau = 1$, the sum of all these contributions equals the total risk of the portfolio $\mathbf{m}$. The ratio $\frac{m_i}{\rho(\mathbf{m})}\frac{\partial\rho(\mathbf{m})}{\partial m_i}$ tells us the percentage of total risk contributed by asset i. The term $\frac{\partial\rho(\mathbf{m})}{\partial m_i}$ is referred to as *marginal risk* and represents the impact on the overall risk from a small change in the position size of asset i, keeping all other positions fixed. If the sign of the marginal risk of asset i is positive then investing a small amount of money more in asset i will increase the portfolio risk; and if the sign is negative we have the opposite effect. This means that the assets with negative marginal risk can be used as hedging instruments and for diversification. Note that all these considerations make sense only if the increase ϵ in asset i is small enough so that the *incremental risk* can be aproximated by

$$\rho(\mathbf{m} + \epsilon\mathbf{e}_i) - \rho(\mathbf{m}) \approx \epsilon\frac{\partial\rho(\mathbf{m})}{\partial m_i}, \tag{1.3.6}$$

where e_i denotes the canonical basis vector of $\mathbb{R}^n$. For example, if the risk contribution of asset 1, in a portfolio consisting of two assets, is twice that of asset 2, the above relation does not imply that eliminating asset 1 from the portfolio will reduce the total risk by $\frac{2}{3}$. This is because the partial derivatives describe the risk changes only locally and both the marginal risk and the total risk will change as the position in the asset changes. At this point a few questions may seem natural. Is each of the term in the decomposition (1.3.2) really an appropriate representation of the risk contribution of each individual asset? If yes, is (1.3.2) the only representation that makes sense, or are there other plausible ways of decomposing the total risk? These questions have been answered by Tasche in [24]: he has postulated a general assumption of suitability of the risk contribution, and shown that the only appropriate representation of the marginal risk is the first order partial derivative of the risk measure with respect to the position size. The approach by Tasche [24] is very instructive and provides intuition about the problem, we want therefore to discuss it. This approach is an axiomatic one, we first define the evident properties that risk contributions must satisfy and then look for possible candidates. We start defining the *Return on Risk-Adjusted-Capital*, a well-known concept in economy in the context of capital allocation,

$$RORAC(\mathbf{m}) = \frac{E[\mathbf{m}'\mathbf{R}]}{\rho(\mathbf{m}) - E[\mathbf{m}'\mathbf{R}]}. \tag{1.3.7}$$

The RORAC of a portfolio m is defined as ration between the expected profit of the portfolio $E[m'R]$ and the *economic capital* $\rho(m) - E[m'R]$, which is the amount of capital needed to cover the risks at some confidence level. The *per-unit RORAC of asset* i is given by

$$\frac{E[R_i]}{\rho(R_i) - E[R_i]}, \tag{1.3.8}$$

where $\rho(R_i)$ is the risk of asset i per unit of position. The RORAC is used to measure the performance of the portfolio: if the RORAC of portfolio A is higher than that of portfolio B, then portfolio A gives a higher return per unit of risk than portfolio B, and so has a better performance than B in this terminology. Observe that the case of opposite signs in the numerator and denominator respectively of the RORAC is unrealistic. The case of positive numerator and negative economic capital means that we can invest with positive expected profit and we are even paid to do so. The opposite case means that we are paying for being allowed to bear someone else's risk. Mathematically, we can describe the problem of finding risk contributions as finding a vector field $\mathbf{a} = (a_1, \ldots, a_n) : M \to \mathbb{R}^n$ satisfying certain properties.

Definition 1.3.2. *Let $\rho : M \to \mathbb{R}$ be a risk measure on a set of portfolios M. A vector field $\mathbf{a} = (a_1, \ldots, a_d) : M \to \mathbb{R}^n$ is said to be suitable for performance measurement for ρ if it satisfies the following conditions.*

1. *For all $\mathbf{m} \in M$ and $i \in \{1, \ldots, n\}$ the inequality*

$$\frac{E[R_i]}{a_i(\mathbf{m}) - E[R_i]} > \frac{E[\mathbf{m}'\mathbf{R}]}{\rho(\mathbf{m}) - E[\mathbf{m}'\mathbf{R}]}, \tag{1.3.9}$$

implies that there is an $\epsilon > 0$ such that for all $t \in (0, \epsilon)$ we have

$$\begin{aligned} \frac{E[\mathbf{m}'\mathbf{R} + tR_i]}{\rho(\mathbf{m} + t\mathbf{e}_i) - E[\mathbf{m}'\mathbf{R} + tR_i]} &> \frac{E[\mathbf{m}'\mathbf{R}]}{\rho(\mathbf{m}) - E[\mathbf{m}'\mathbf{R}]} \\ &> \frac{E[\mathbf{m}'\mathbf{R} - tR_i]}{\rho(\mathbf{m} - t\mathbf{e}_i) - E[\mathbf{m}'\mathbf{R} - tR_i]}. \end{aligned} \tag{1.3.10}$$

2. *For all* $m \in M$ *and* $i \in \{1, \ldots, n\}$ *the inequality*

$$\frac{E[R_i]}{a_i(\mathbf{m}) - E[R_i]} < \frac{E[\mathbf{m}'\mathbf{R}]}{\rho(\mathbf{m}) - E[\mathbf{m}'\mathbf{R}]}, \tag{1.3.11}$$

implies that there is an $\epsilon > 0$ *such that for all* $t \in (0, \epsilon)$ *we have*

$$\begin{aligned} \frac{E[\mathbf{m}'\mathbf{R} + tR_i]}{\rho(\mathbf{m} + t\mathbf{e}_i) - E[\mathbf{m}'\mathbf{R} + tR_i]} &< \frac{E[\mathbf{m}'\mathbf{R}]}{\rho(\mathbf{m}) - E[\mathbf{m}'\mathbf{R}]} \\ &< \frac{E[\mathbf{m}'\mathbf{R} - tR_i]}{\rho(\mathbf{m} - t\mathbf{e}_i) - E[\mathbf{m}'\mathbf{R} - tR_i]}. \end{aligned} \tag{1.3.12}$$

To better understand why 1 and 2 are two universal properties that the risk contributions must satisfy in order to be coherent with their interpretation, we explain the meaning of the above conditions. The inequality (1.3.9) tells us that the performance of one unit of the asset i as part of the portfolio is higher than that of the entire portfolio. If such a relation holds, the inequalities (1.3.10) tells us that investing a little more (less) in asset i should increase (decrease) the performance of the entire portfolio. The inequalities (1.3.11) and (1.3.12) are the corresponding ones with reverted inequality signs. Note that the initial risk measure $\rho(R_i)$ in the per-unit individual RORAC of asset i is replaced by $a_i(\mathbf{m})$, the candidate for the risk contribution measure of asset i in the portfolio. Tasche [24] shows that under certain assumptions the only candidate for the risk contribution vector field satisfying the above suitability properties based on RORAC is the vector of the partial derivatives. Here the result exposed formally.

Theorem 1.3.2. *Let* $\rho : M \to \mathbb{R}$ *be a partially differentiable risk measure on an open set of portfolios* M *with continuous partial derivatives. Let* $\mathbf{a} = (a_1, \ldots, a_n) : M \to \mathbb{R}^n$ *be a continuous vector field. Then* $\mathbf{a}$ *is suitable for performance measurement with* ρ *if and only if*

$$a_i(\mathbf{m}) = \frac{\partial \rho}{\partial m_i}(\mathbf{m}), \quad \text{for all} \;\; i = 1, \ldots, n, \quad \mathbf{m} \in M \tag{1.3.13}$$

Proof. We explain the argument proposed by Tasche [24]. For $\mathbf{m} \in M$ with $\rho(\mathbf{m}) \neq E[\mathbf{m}'\mathbf{R}]$, and $i = 1, \ldots, n$ it holds

$$\begin{aligned} \frac{\partial RORAC(\mathbf{m})}{\partial m_i} &= \frac{1}{\big(\rho(\mathbf{m}) - E[\mathbf{m}'\mathbf{R}]\big)^2} \left(\rho(\mathbf{m}) E[R_i] - \frac{\partial \rho}{\partial m_i} E[\mathbf{m}'\mathbf{R}]\right) \\ &= \frac{1}{\big(\rho(\mathbf{m}) - E[\mathbf{m}'\mathbf{R}]\big)^2} \Bigg(\rho(\mathbf{m}) E[R_i] - a_i(\mathbf{m}) E[\mathbf{m}'\mathbf{R}] \\ &\quad + \left(a_i(\mathbf{m}) - \frac{\partial \rho}{\partial m_i}(\mathbf{m})\right) E[\mathbf{m}'\mathbf{R}]\Bigg). \end{aligned} \tag{1.3.14}$$

If (1.3.13) is satisfied then for any $i = 1, \dots, n$ condition 1 (or 2) in Definition 1.3.2 implies $\frac{\partial RORAC(\mathbf{m})}{\partial m_i} > 0 (< 0)$ and therefore (1.3.10) and (1.3.12) are satisfied. For the other direction fix any $i = 1, \dots, n$ and note that by continuity we only need to show (1.3.13) for $\mathbf{m} \in M$ such that $m_i \neq 0$ and $m_j \neq 0$ for some $j \neq i$. We consider separately the following cases.

1. $a_i(\mathbf{m}) \neq 0$, $\rho(\mathbf{m}) \neq 0$, $\rho(\mathbf{m}) \neq m_i a_i(\mathbf{m})$.
2. $a_i(\mathbf{m}) \neq 0$, $\rho(\mathbf{m}) \neq 0$, $\rho(\mathbf{m}) = m_i a_i(\mathbf{m})$.
3. $a_i(\mathbf{m}) = 0$, $\rho(\mathbf{m}) \neq 0$.
4. $\rho(\mathbf{m}) = 0$, each neighborhood of $\mathbf{m}$ contains some $\mathbf{q}$ such that $\rho(\mathbf{q}) \neq 0$.
5. $\rho(\mathbf{q}) = 0$, for all $\mathbf{q}$ is some neighborhood of m.

Note that 5 is clear since in this case $\rho = 0$ and $\frac{\partial RORAC(\mathbf{m})}{\partial m_i} = 0$ in some neighborhood of $\mathbf{m}$, and 4 follows by continuity once we have proved the statement in the other three cases. For case 1 choose the returns $\mathbf{R}(t)$ such that

$$E[R_k(t)] = \begin{cases} 1 & k = i \\ \frac{t}{m_j}\left(\frac{\rho(\mathbf{m})}{a_i(\mathbf{m})} - m_i\right) & k = j \\ 0 & k \neq i, j \end{cases} \tag{1.3.15}$$

for $t \in (0,1)$. For these expected returns we have

$$\begin{aligned} \mathbf{m}' E[\mathbf{R}(t)] &= t \frac{\rho(\mathbf{m})}{a_i(\mathbf{m})} + (1-t) m_i \\ E[R_i(t)]\rho(\mathbf{m}) - a_i(\mathbf{m})\mathbf{m}' E[\mathbf{R}(t)] &= (1-t)(\rho(\mathbf{m}) - m_i a_i(\mathbf{m})). \end{aligned} \tag{1.3.16}$$

Thus, it is possible to choose two sequences $(t_k), (s_k)$ with $t_k, s_k \to 1$, such that $\mathbf{m}' E[\mathbf{R}(s_k)] \neq \rho(\mathbf{m}) \neq \mathbf{m}' E[\mathbf{R}(t_k)]$, (t_k) satisfies (1.3.9) and (s_k) satisfies (1.3.11). By suitableness and (1.3.14) we have

$$\begin{aligned} &(1-t_k)(\rho(\mathbf{m}) - m_i a_i(\mathbf{m})) \\ &+ \left(a_i(\mathbf{m}) - \frac{\partial \rho}{\partial m_i}(\mathbf{m})\right)\left(t_k \frac{\rho(\mathbf{m})}{a_i(\mathbf{m})} + (1-t_k) m_i\right) \geq 0, \\ &(1-s_k)(\rho(\mathbf{m}) - m_i a_i(\mathbf{m})) \\ &+ \left(a_i(\mathbf{m}) - \frac{\partial \rho}{\partial m_i}(\mathbf{m})\right)\left(s_k \frac{\rho(\mathbf{m})}{a_i(\mathbf{m})} + (1-s_k) m_i\right) \leq 0. \end{aligned} \tag{1.3.17}$$

Now $k \to \infty$ yields (1.3.13). For case 2 or 3 proceed analogously. □

The main limitation of this approach is the use of RORAC for performance measurement, which is appropriate for banks, but not necessarily in every financial context, thus the above argument may not be valid in general. However, the use of the partial derivatives can be justified by the fact that it describes how the total risk of the portfolio changes if there is a small local change in one of the positions. Another justification, based on the notion of "fairness" in cooperative game theory, was given by Denault [5].

1.4 Differentiating Risk Measures

Considering the above discussion we are now interested in examining if the most common risk measures can be differentiated and in computing these partial derivatives. We also want to see if under the most common distribution assumptions for the portfolio returns we can explicitly find these derivatives. This is interesting because we would like to use the information given by the risk contributions to construct risk-balanced portfolios. We treat systematically the differentiation of the risk measures considered in Section 1.2.

1.4.1 Derivatives of the Standard Deviation and Covariance Based Risk Contributions

For $\mathbf{m} \in M \subset \mathbb{R}^n$ we have

$$\rho(\mathbf{m}) = \sqrt{\mathbf{m}'\Omega\mathbf{m}}, \tag{1.4.1}$$

where we use the same notation as in (1.2.2). The partial derivatives are given by

$$\begin{aligned} \frac{\partial\rho(\mathbf{m})}{\partial m_k} &= \frac{1}{2\sqrt{\mathbf{m}'\Omega\mathbf{m}}}\frac{\partial}{\partial m_k}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} m_i m_j \Omega_{ij}\right) \\ &= \frac{1}{\sqrt{\mathbf{m}'\Omega\mathbf{m}}}\sum_{j=1}^{n} m_j\Omega_{kj} \\ &= \frac{1}{\sqrt{\mathbf{m}'\Omega\mathbf{m}}}(\Omega\mathbf{m})_k. \end{aligned} \tag{1.4.2}$$

The risk contributions are therefore given by

$$\frac{m_i}{\sqrt{\mathbf{m}'\Omega\mathbf{m}}}(\Omega\mathbf{m})_i, \quad i = 1, \dots, d \tag{1.4.3}$$

and depend only on the covariances $Cov(R_i, R_j)$ for $i \neq j$ and variances $Var[R_i]$ of the asset returns. The problem of estimating these parameters will be considered later. The decomposition of the homogeneous measure σ can be written as

$$\sigma(\mathbf{m}) = \frac{1}{\sigma(\mathbf{m})} \sum_{i=1}^{n} m_i \mathbf{e}_i' \Omega \mathbf{m}, \tag{1.4.4}$$

where $\mathbf{e}_i$ is the standard basis vector in $\mathbb{R}^n$. We now want to compare the risk contributions obtained with partial derivatives with the covariance based risk contributions widely used in practice. The covariance based risk contributions are obtained using the notion of *best linear predictor* for random variables in $\mathcal{L}^2$.

Definition 1.4.1. *Let Y, Z be two square integrable random variables so that $Var[Z] > 0$. For $z \in \mathbb{R}$ we define the projection $\pi_Z(z, Y)$ of $Y - E[Y]$ onto the linear space spanned by $Z - E[Z]$ via*

$$\pi_Z(z, Y) = \frac{Cov(Y, Z)}{Var[Z]} z. \tag{1.4.5}$$

The function $\pi_Z(z, Y)$ is the best linear predictor of $Y - E[Y]$ given $Z - E[Z] = z$ in the sense that the random variable $\pi_Z\big(Z - E[Z], Y\big)$ minimizes the $\mathcal{L}^2$-distance between $Y - E[Y]$ and the linear space spanned by $Z - E[Z]$. Indeed, for $\theta \in \mathbb{R}$,

$$\begin{aligned} &E\Big[\big((Y - E[Y]) - \theta(Z - E[Z])\big)^2\Big] = Var[Y - \theta Z] = \\ &= Var[Y] - 2\theta Cov(Y, Z) + \theta^2 Var[Z] \end{aligned} \tag{1.4.6}$$

and minimizing over θ shows the statement. This has been applied in the context of risk contributions in the following way: we define the set of portfolios $M \subset \mathbb{R}^n$ as

$$M := \big\{\mathbf{m} \in \mathbb{R}^n : Var[\mathbf{m}'\mathbf{R}] > 0\big\}, \tag{1.4.7}$$

and note that for any $\mathbf{m} \in M$, $\lambda > 0$ it holds $\lambda \mathbf{m} \in M$. The idea of this method is to set $Z := \mathbf{m}'\mathbf{R}$, $Y := m_i R_i$ and choose a value for $z = \mathbf{m}'\mathbf{R} - E[\mathbf{m}'\mathbf{R}]$ corresponding to a worst-case scenario and use the best linear predictor to determine the risk contribution. We set $z = \rho(\mathbf{m})$ to be the portfolio risk. The vector field of risk contributions obtained with this method is then defined by

$$\begin{aligned} a_i(\mathbf{m}) &:= \pi_{\mathbf{m}'\mathbf{R}}\big(\rho(\mathbf{m}), m_i R_i\big) = \frac{Cov(m_i R_i, \mathbf{m}'\mathbf{R})}{Var[\mathbf{m}'\mathbf{R}]} \rho(\mathbf{m}) \\ &= \frac{m_i \sum_{j=1}^{n} m_j Cov(R_i, R_j)}{Var[\mathbf{m}'\mathbf{R}]} \rho(\mathbf{m}) = \frac{m_i (\Omega \mathbf{m})_i}{\mathbf{m}'\Omega\mathbf{m}} \rho(\mathbf{m}), \end{aligned} \tag{1.4.8}$$

where Ω denotes the covariance matrix of $\mathbf{R}$. We interpret the so constructed $a_i(\mathbf{m})$ as the best linear predictor of the profit fluctuation of asset i given that the portfolio profit fluctuation is just the risk $\rho(\mathbf{m})$. For the standard deviation $\rho(\mathbf{m}) = \sigma(\mathbf{m})$ we get

$$a_i(\mathbf{m}) = \frac{m_i(\Omega\mathbf{m})_i}{\sqrt{\mathbf{m}'\Omega\mathbf{m}}} = m_i \frac{\partial\sigma(\mathbf{m})}{\partial m_i}, \tag{1.4.9}$$

i.e. in this case we have the same risk contribution as the ones obtained with partial derivatives and since $\sigma : M \to \mathbb{R}$ is homogenous we can decompose the total risk in the risk contributions as in (1.3.2). However, we are going to see that for other risk measures the a_i's obtained using the best linear predictor can differ considerably from the partial derivatives and should not be used for risk contributions measurements even if

$$\rho(\mathbf{m}) = \sum_{i=1}^{n} m_i a_i(\mathbf{m}) \tag{1.4.10}$$

still holds.

1.4.2 Derivatives of the Value-At-Risk

We have a closer look at the problem of differentiating VaR, although, as explained in Section 1.2, VaR represents a problematic risk measure. This is anyway interesting because we are going to deal with the problem of differentiating the lower α-quintile of the profit distribution, which is relevant also when considering the expected shortfall derivatives. In general the quantile function $q_\alpha(m)$ is not differentiable in m. In order to guarantee the existence of the partial derivatives we need to impose some technical assumptions on the distribution of the random vector $\mathbf{R} = (R_1, \ldots, R_n)$. Tasche [25] discusses this problem in mathematical details. We expose the essential theoretical aspects treated in [25]. We begin recalling a basic definition from probability theory.

Definition 1.4.2. *For the random vector $\mathbf{R} = (R_1, \ldots, R_n)$, R_1 is said to have a conditional density given $(R_2, \ldots, R_n)$ if it exists a measurable function $\gamma : \mathbb{R}^n \to [0, \infty)$ such that for all $A \in \mathcal{B}(\mathbb{R})$*

$$P[R_1 \in A | R_2, \ldots, R_n] = \int_A \gamma(t, R_2, \ldots, R_n) dt. \tag{1.4.11}$$

Note that the existence of a joint density of $\mathbf{R}$ implies the existence of a conditional density but not necessarily vice versa. Very useful for the computation of the quantile derivatives is the following result.

Lemma 1.4.1. *Assume that R_1 has a conditional density given $(R_2, \ldots, R_n)$. Then for any portfolio $\mathbf{m}$ with $m_1 \neq 0$ we have*

1. *The random variable* $X = \mathbf{m}'\mathbf{R}$ *has a density given by*

$$f_X(x) = \frac{1}{|m_1|} E\Big[\gamma\Big(\frac{x - \sum_{i=2}^n m_i R_i}{m_1}, R_2, \ldots, R_n\Big)\Big]. \tag{1.4.12}$$

2. *If* $f_X(x) > 0$ *we have for* $i = 2, \ldots, n$,

$$E[R_i|X = x] = \frac{E\Big[R_i \gamma\Big(\frac{1}{m_1}\big(x - \sum_{j=2}^n m_j R_j\big), R_2, \ldots, R_n\Big)\Big]}{E\Big[\gamma\Big(\frac{1}{m_1}\big(x - \sum_{j=2}^n m_j R_j\big), R_2, \ldots, R_n\Big)\Big]}. \tag{1.4.13}$$

3. *If* $f_X(x) > 0$ *we have*

$$E[R_1|X = x] = \frac{E\Big[\frac{x - \sum_{j=2}^n m_j R_j}{m_1} \gamma\Big(\frac{1}{m_1}\big(x - \sum_{j=2}^n m_j R_j\big), R_2, \ldots, R_n\Big)\Big]}{E\Big[\gamma\Big(\frac{1}{m_1}\big(x - \sum_{j=2}^n m_j R_j\big), R_2, \ldots, R_n\Big)\Big]}. \tag{1.4.14}$$

Proof. 1. Consider $m_1 > 0$ (for $m_1 < 0$ proceed analogue), using the property of conditional expectations we have

$$\begin{aligned} P[X \le x] &= E[1_{\{X \le x\}}] = E\big[E[1_{\{X \le x\}}|R_2, \ldots, R_n]\big] \\ &= E\big[E[1_{\{R_1 \le m_1^{-1}(x - \sum_{i=2}^n m_i R_i)\}}|R_2, \ldots, R_n]\big] \\ &= E\Big[\int_{-\infty}^{m_1^{-1}(x - \sum_{i=2}^n m_i R_i)} \gamma(u, R_2, \ldots, R_n) du\Big] \\ &= E\Big[\int_{-\infty}^{x} \frac{1}{m_1} \gamma\Big(\frac{u - \sum_{i=2}^n m_i R_i}{m_1}, R_2, \ldots, R_n\Big) du\Big] \\ &= \int_{-\infty}^{x} E\Big[\frac{1}{m_1} \gamma\Big(\frac{u - \sum_{i=2}^n m_i R_i}{m_1}, R_2, \ldots, R_n\Big)\Big] du, \end{aligned} \tag{1.4.15}$$

where in the last step we have exchanged the order of integration, which is allowed by Fubini's theorem. Therefore the density of X is given by the integrand in the right hand side.

2. Provided that $f_X(x) > 0$ we can write

$$\begin{aligned} E[R_i|X = x] &= \frac{E[R_i 1_{\{X = x\}}]}{P[X = x]} = \lim_{\delta \downarrow 0} \frac{\delta^{-1} E[R_i 1_{\{x < X \le x + \delta\}}]}{\delta^{-1} P(x < X \le x + \delta)} \\ &= \frac{\frac{\partial}{\partial x} E[R_i 1_{\{X \le x\}}]}{f_X(x)}. \end{aligned} \tag{1.4.16}$$

Note that exchanging the limit and the expectation can be justified by dominated convergence. The denominator was calculated in 1 and the numerator can be computed as in 1, for $m_1 > 0$ and $i = 2, \dots, n$,

$$\begin{aligned} \frac{\partial E[R_i 1_{\{X \le x\}}]}{\partial x} &= \frac{\partial}{\partial x} E\big[E[R_i 1_{\{X \le x\}} | R_2, \dots, R_n]\big] = \\ &= \frac{\partial}{\partial x} E\big[R_i E[1_{\{X \le x\}} | R_2, \dots, R_n]\big] = \\ &= \frac{\partial}{\partial x} E\Big[R_i \int_{-\infty}^{x} \frac{1}{m_1} \gamma\Big(\frac{u - \sum_{i=2}^n m_i R_i}{m_1}, R_2, \dots, R_n\Big) du\Big] = \\ &= \frac{1}{m_1} E\Big[R_i \gamma\Big(\frac{x - \sum_{i=2}^n m_i R_i}{m_1}, R_2, \dots, R_n\Big)\Big], \end{aligned} \tag{1.4.17}$$

where we have exchanged again integration w.r.t. u and expectation, and then taken the derivative with respect to x. The identity 2 follows form 1 and the last equality.

3. We can write

$$E[R_1 | X = x] = E\Big[\frac{x - \sum_{j=2}^n m_j R_j}{m_1} \Big| X = x\Big], \tag{1.4.18}$$

and applying 2 we get 3.

$\square$

The quantities determined in the above result motivate the following assumptions on γ made by Tasche in [25].

Assumptions 1.4. Let the conditional density γ satisfy the following assumptions.

1. For fixed $r_2, \dots, r_n$ the map $t \mapsto \gamma(t, r_2, \dots, r_n)$ is continuous in t.
2. The map $(t, m) \mapsto E\Big[\gamma\Big(\frac{x - \sum_{i=2}^n m_i R_i}{m_1}, R_2, \dots, R_n\Big)\Big]$ is finite valued and continuous.
3. For $i = 2, \dots, n$ the map $(t, m) \mapsto E\Big[R_i \gamma\Big(\frac{x - \sum_{i=2}^n m_i R_i}{m_1}, R_2, \dots, R_n\Big)\Big]$ is finite valued and continuous.

Typical situations where these assumptions are satisfied are listed by Tasche [25] in Remark 2.4. Tasche [25] gives in Lemma 3.2 and Theorem 3.3 rigorous mathematical arguments, using the implicit function theorem, showing that if the above assumptions are satisfied then the quantile function q_α is partially differentiable. Once we know that q_α is differentiable with respect to the components of m we can obtain representations of the partial derivatives in terms of expectations of conditional distributions by applying Lemma 1.4.1.

Theorem 1.4.1. *Assume that the distribution of the returns is such that there exists a conditional density of R_1 given $R_2, \ldots, R_n$ satisfying Assumptions 1.4 in some open set $M \subset \mathbb{R} \setminus \{0\} \times \mathbb{R}^{d-1}$, and that $f_X(q_\alpha(\mathbf{m})) > 0$. Then, q_α is partially differentiable at m with*

$$\frac{\partial q_\alpha}{\partial m_i}(\mathbf{m}) = E\left[R_i \middle| \sum_{i=1}^{n} m_i R_i = q_\alpha(\mathbf{m})\right] \tag{1.4.19}$$

Proof. By Lemma 1.4.1 the random variable $X = \sum_{i=1}^n m_i R_i$ is continuous with density function given by (1.4.12) and so we have for $m_1 > 0$ by (1.4.15)

$$\alpha = P[X \le q_\alpha(X)] = E\left[\int_{-\infty}^{m_1^{-1}(q_\alpha(X) - \sum_{i=2}^n m_i R_i)} \gamma(u, R_2, \ldots, R_n) du\right]. \tag{1.4.20}$$

We can now try to differentiate the above equation on both sides. By [6] Theorem A.9.1 we can differentiate under the expectation. Differentiating with respect to m_i for $i = 2, \ldots, n$, we have

$$0 = \frac{1}{m_1} E\left[\Big(\frac{\partial q_\alpha(X)}{\partial m_i} - R_i\Big)\gamma\Big(m_1^{-1}\Big(q_\alpha(X) - \sum_{i=2}^{n} m_i R_i\Big), R_2, \ldots, R_n\Big)\right]. \tag{1.4.21}$$

Solving the above equation for $\frac{\partial q_\alpha(X)}{\partial m_i}$ and using Lemma 1.4.1 2 we find the desired equality for $i = 2, \ldots, n$. Note that by assumption we can divide by

$$\frac{1}{m_1} E\left[\gamma\Big(m_1^{-1}\Big(q_\alpha(X) - \sum_{i=2}^{n} m_i R_i\Big), R_2, \ldots, R_n\Big)\right] = f_X(q_\alpha(m)) > 0. \tag{1.4.22}$$

Differentiating (1.4.20) with respect to m_1 yields

$$\begin{aligned} 0 = E\bigg[\Big(&-\frac{1}{m_1^2}\Big(q_\alpha(m) - \sum_{i=2}^{n} m_i R_i\Big) + \frac{1}{m_1}\frac{\partial q_\alpha(X)}{\partial m_1}\Big) \times \\ &\times \gamma\Big(m_1^{-1}\Big(q_\alpha(X) - \sum_{i=2}^{n} m_i R_i\Big), R_2, \ldots, R_n\Big)\bigg] \end{aligned} \tag{1.4.23}$$

and solving with respect to m_1 and applying Lemma 1.4.1 3 we find the desired expression for the partial derivative with respect to m_1. □

Under the above assumptions we have for the Value-at-Risk,

$$\frac{\partial VaR_\alpha}{\partial m_i}(\mathbf{m}) = -E\left[R_i \middle| -\sum_{i=1}^{n} m_i R_i = VaR_\alpha(\mathbf{m})\right] \tag{1.4.24}$$

and the risk decomposition into individual risk contributions is given by,

$$VaR_\alpha(\mathbf{m}) = -\sum_{i=1}^{n} m_i E\Big[R_i\Big| - X = VaR_\alpha(\mathbf{m})\Big]. \tag{1.4.25}$$

We consider again the simple situations of Examples 1.1.1 and 1.1.4.

Example 1.4.1. Using the explicit formulas of VaR given in Example 1.2.2 we can compute the partial derivatives. For the multivariate normal model,

$$\frac{\partial VaR_\alpha}{\partial m_i}(\mathbf{m}) = -\frac{(\Omega\mathbf{m})_i}{\sqrt{\mathbf{m}'\Omega\mathbf{m}}}\Phi^{-1}(\alpha) - \mu_i, \tag{1.4.26}$$

and for the elliptical model of 1.1.4

$$\frac{\partial VaR_\alpha}{\partial m_i}(\mathbf{m}) = -\frac{(\Omega\mathbf{m})_i}{\sqrt{\mathbf{m}'\Omega\mathbf{m}}}T_\nu^{-1}(\alpha) - \mu_i. \tag{1.4.27}$$

The Euler decomposition of the positive homogeneous measure VaR is then given by

$$VaR_\alpha(\mathbf{m}) = \sum_{i=1}^{n} m_i\Big(- \frac{(\Omega\mathbf{m})_i}{\sqrt{\mathbf{m}'\Omega\mathbf{m}}}\Phi^{-1}(\alpha) - \mu_i\Big). \tag{1.4.28}$$

and analogously for the other model. We can observe that the risk contributions obtained by differentiating VaR differ from the ones obtained by the best linear predictor:

$$\frac{m_i(\Omega\mathbf{m})_i}{\mathbf{m}'\Omega\mathbf{m}}VaR_\alpha(\mathbf{m}) = -\frac{m_i(\Omega\mathbf{m})_i}{\sqrt{\mathbf{m}'\Omega\mathbf{m}}}\Phi^{-1}(\alpha) - \frac{m_i(\Omega\mathbf{m})_i}{\mathbf{m}'\Omega\mathbf{m}}\mathbf{m}'\boldsymbol{\mu}. \tag{1.4.29}$$

1.4.3 Derivatives of the Expected Shortfall

Once we have done all the work of Section 1.4.2 considering the partial derivatives of the lower quantile function, it is quite clear how to find analogue representations of the partial derivatives for the expected shortfall.

Theorem 1.4.2. *Assume that the distribution of* $\mathbf{R}$ *satisfies the assumptions of Theorem 1.4.1. Then the risk measure* ES_α *is partially differentiable with partial derivatives given by*

$$\frac{\partial ES_\alpha(X)}{\partial m_i} = -\frac{1}{\alpha}\Big\{E\big[R_i 1_{\{X\le q_\alpha(X)\}}\big] + E\big[R_i|X = q_\alpha(X)\big]\big(\alpha - P[X \le q_\alpha(X)]\big)\Big\} \tag{1.4.30}$$

for any $X \in \mathcal{M}$ *with* $E[X^-] < \infty$.

Proof. We use the representation (1.2.29) and differentiate under the integral,

$$\begin{aligned}\frac{\partial ES_\alpha(X)}{\partial m_i} &= \frac{1}{\alpha}\int_0^\alpha \frac{\partial VaR_u(X)}{\partial m_i}du \\ &= -\frac{1}{\alpha}\int_0^\alpha E\big[R_i| - X = VaR_u(X)\big]du \\ &= -\frac{1}{\alpha}\int_0^\alpha E\big[R_i|X = q_u(X)\big]du.\end{aligned} \tag{1.4.31}$$

Note that using exactly the same argument as in the proof of Theorem 1.2.3, we can write for any real measurable function f

$$\int_0^\alpha f(q_u(X))du = E\big[f(X)1_{\{X\leq q_\alpha(X)\}}\big] - f(q_\alpha(X))\big(\alpha - P[X \leq q_\alpha(X)]\big). \tag{1.4.32}$$

Appling this on $f(x) := E[R_i|X = x]$ we find

$$\begin{aligned}\int_0^\alpha E[R_i|X = q_u(X)]du = E\big[E[R_i|X]1_{X\leq q_\alpha(X)}\big]+ \\ + E\big[R_i|X = q_\alpha(X)\big]\big(\alpha - P[X \leq q_\alpha(X)]\big).\end{aligned} \tag{1.4.33}$$

Using properties of conditional expectations, the first term on the right hand side can be simplified to

$$\begin{aligned}E\big[E[R_i|X]1_{\{X\leq q_\alpha(X)\}}\big] &= E\big[E[R_i|X]1_{\{X\leq q_\alpha(X)\}}\big] \\ &= E\Big[E\big[R_i 1_{\{X\leq q_\alpha(X)\}}\big|X\big]\Big] \\ &= E\big[R_i 1_{\{X\leq q_\alpha(X)\}}\big],\end{aligned} \tag{1.4.34}$$

which together with the above equations leads to the desired representation of the partial derivatives. □

We compute the risk contributions for the same two distributions.

Example 1.4.2. Using the explicit formulas of VaR given in Example 1.2.4 we can compute the partial derivatives. For the multivariate normal model,

$$\frac{\partial ES_\alpha}{\partial m_i}(\mathbf{m}) = \frac{(\Omega\mathbf{m})_i}{\alpha\sqrt{\mathbf{m}'\Omega\mathbf{m}}}\varphi\big(\Phi^{-1}(\alpha)\big) - \mu_i \tag{1.4.35}$$

and for the elliptic model of 1.1.4,

$$\frac{\partial ES_\alpha}{\partial m_i}(\mathbf{m}) = -\frac{\nu}{1-\nu}\frac{(\Omega\mathbf{m})_i}{\alpha\sqrt{\mathbf{m}'\Omega\mathbf{m}}}\left(1+\frac{\big(T_\nu^{-1}(\alpha)\big)^2}{\nu}\right)t_\nu\big(T_\nu^{-1}(\alpha)\big) - \mu_i. \tag{1.4.36}$$

The similarity of the expressions for the risk contributions in these two different models suggest that by elliptical distributions the risk contributions always "look like the same". In fact, the next result shows that by elliptical distributions, independently on the risk measure under consideration, we can compute risk contributions by differentiation always in the same way ([17] Corollary 6.27)

Theorem 1.4.3. *Let* $\mathbf{R} \sim E_n(\boldsymbol{\mu}, \Omega, \psi)$ *and let* $\rho : M \to \mathbb{R}$ *be a positive homogeneous, translation invariant and partially differentiable risk measure depending only on the distribution of the profit. Then,*

$$\frac{\partial \rho}{\partial m_i}(\mathbf{m}) = \frac{(\Omega \mathbf{m})_i}{\sqrt{\mathbf{m}'\Omega\mathbf{m}}}\rho(Y_1) - \mu_i, \tag{1.4.37}$$

where $\mathbf{R} = A\mathbf{Y} + \boldsymbol{\mu}$ *with* $A \in \mathbb{R}^{n\times k}$ *s.t.* $A'A = \Omega$, $\mathbf{Y} = (Y_1, \ldots, Y_k) \sim S_k(\psi)$ *and* $\boldsymbol{\mu} \in \mathbb{R}^n$.

Proof. By (1.2.19) we have

$$\mathbf{m}'\mathbf{R} \overset{(d)}{\sim} |\mathbf{m}'A|Y_1 + \mathbf{m}'\boldsymbol{\mu}. \tag{1.4.38}$$

Therefore by positive homogeneity and translation invariance we have

$$\begin{aligned}\rho(\mathbf{m}'\mathbf{R}) &= \rho(|\mathbf{m}'A|Y_1 + \mathbf{m}'\boldsymbol{\mu}) = |\mathbf{m}'A|\rho(Y_1) - \mathbf{m}'\boldsymbol{\mu} \\ &= \sqrt{\mathbf{m}'A(\mathbf{m}'A)'}\rho(Y_1) - \mathbf{m}'\boldsymbol{\mu} = \sqrt{\mathbf{m}'\Omega\mathbf{m}}\rho(Y_1) - \mathbf{m}'\boldsymbol{\mu},\end{aligned} \tag{1.4.39}$$

and taking partial derivatives of the right hand side proves the statement. □

When we are dealing with elliptical returns models we can therefore compute the risk contributions explicitly and we do not have to rely on the representations (1.4.24) and (1.4.30). We are going to use these two formulas to estimate the risk contributions under other distributions assumptions where explicit formulas are not available.

Chapter 2

Portfolio Construction

2.1 Classical Methods and Risk Contributions

We begin this chapter with a brief outline of the classical portfolio selection theory and discuss the drawbacks in practical applications of the classical solutions. The aim of this section is not to discuss every aspect of the classical portfolio theory in detail, but rather to motivate the necessity of a new approach that takes into account the risk contributions we have introduced in the previous chapter. A complete description of the classical portfolio selection theory can be found in [18] Chapter 6 and [10]. We now introduce a few additional notation in order to formulate the selection problem in terms of *portfolio weights*. We fix a certain initial wealth $C > 0$ the investor wants to allocate in the n investment opportunities. For a certain portfolio $\mathbf{m}$ we define the vector of portfolio weights $\mathbf{w} = (w_1, \ldots, w_n)$ by

$$w_i = \frac{m_i}{C}, \tag{2.1.1}$$

for $i = 1 \ldots, n$. The i-th component of this vector represents the fraction of the initial capital invested in instrument i. In this way, we can represent a portfolio with a vector $\mathbf{w} \in \mathbb{R}^n$ of portfolio weights and a set of admissible portfolios with a corresponding set of portfolio weights W. Using this notation the *portfolio return* is given by $\mathbf{w}'\mathbf{R}$. Typical constraints on the set of admissible strategies are

1. $W_1^+ = \{\mathbf{w} \in [0,1]^n : \mathbf{1}'\mathbf{w} = 1\}$
2. $W_l^+ = \{\mathbf{w} \in \mathbb{R}_+^n : \mathbf{1}'\mathbf{w} = l\}$ for a certain leverage $l \geq 1$.
3. $W_l = \{\mathbf{w} \in \mathbb{R}^n : \mathbf{1}'\mathbf{w} = l\}$ for a certain $l \geq 1$.

In the first case we do not allow borrowing of additional capital to invest or short selling of instruments. In the second case we can borrow additional capital at leverage l, in the third one, borrowing and short-selling are both allowed. It is convenient for notational purposes when considering the problem of portfolio selection to characterize strategies using weights instead of amounts of money as we did in Chapter 1. Fix now a certain initial wealth $C > 0$ and a certain set of admissible strategies W. Let ρ be a partially differentiable positive homogenous risk measure defined on an open set that contains all vectors of the form $\lambda\mathbf{w}$ for $\mathbf{w} \in W$ and $\lambda > 0$ (cone). In this setting the portfolio risk is proportional to the quantity $\rho(\mathbf{w}) = \rho(\mathbf{w}'\mathbf{R})$, which is the risk of the portfolio return and can be interpreted as the portfolio risk per unit of initial capital invested. The problem of portfolio selection consists in formulating a criterion to select an optimal investment strategy $\mathbf{w}^* \in W$. A well-known method is the *mean-ρ* approach, that consists in minimizing the portfolio risk, under a fixed expected portfolio return, over all admissible strategies. Precisely, we select the optimal strategy as follows.

Definition 2.1.1. *The mean-ρ optimal portfolio $\mathbf{w}^* \in W$ for a certain expected return $\mu^* \in \{E[\mathbf{w}'\mathbf{R}] : \mathbf{w} \in W\}$ is defined by*

$$\mathbf{w}^* = \arg\min\nolimits_{\mathbf{w}\in W}\rho(\mathbf{w}) \quad \textit{under the constraint} \quad E[\mathbf{w}'\mathbf{R}] = \mu^*. \tag{2.1.2}$$

This approach has been proposed by Markowitz in [14] and [15] more than fifty years ago using the variance $\rho(\mathbf{w}) = \sigma^2(\mathbf{w}) = \mathbf{w}'\Omega\mathbf{w}$ as risk measure. Markowitz approach is also referred to in the literature as *mean-variance* approach. In case of elliptical distributed return vector $\mathbf{R}$, there is however no difference between minimizing the risk under any positive homogenous translation invariant risk measure and minimizing under the variance. Precisely we have the following result from [17] Proposition 6.13.

Theorem 2.1.1. *Let $\mathbf{R} \sim E_n(\boldsymbol{\mu}, \Omega, \psi)$ with $Var[R_i] < \infty$ for all $i = 1, \dots, n$. Assume that ρ is also translation invariant and depends only on the distribution of the profit. Then,*

$$\arg\min\nolimits_{\mathbf{w}\in W, E[\mathbf{w}'\mathbf{R}]=\mu^*}\rho(\mathbf{w}) = \arg\min\nolimits_{\mathbf{w}\in W, E[\mathbf{w}'\mathbf{R}]=\mu^*}\mathbf{w}'\Omega\mathbf{w}. \tag{2.1.3}$$

Proof. Recall (1.2.19),

$$\mathbf{w}'R \overset{(d)}{\sim} \mathbf{w}'\Omega\mathbf{w}Y_1 + \mathbf{w}'\mu, \tag{2.1.4}$$

form which it follows that for every portfolio $\mathbf{w} \in W$ with $E[\mathbf{w}'\mathbf{R}] = \mu^*$ we have

$$\rho(\mathbf{w}'\mathbf{R}) = \rho\big(\mathbf{w}'\Omega\mathbf{w} + \mu^*\big) = \mathbf{w}'\Omega\mathbf{w}\rho(Y_1) - \mu^*, \tag{2.1.5}$$

which implies the statement. $\square$

The mean-variance approach by Markowitz appears very powerful and elegant. Even if for general constraints there is no analytical solution for the minimization problem, the *quadratic programming* method of numerical optimization can be be applied to minimize efficiently the quadratic function $\mathbf{w}'\Omega\mathbf{w}$ under linear constraints (see [21] Chapter 16). Despite these appealing features, this approach involves several difficulties when applied in practice. First of all, the output of the optimization problem is very sensitive to changes in the input parameters. It has been observed, see for example [4], that small variations in estimating the expected returns of the instruments can lead to significantly different strategies. Secondly, the mean-ρ optimal portfolios optimize only the overall risk of the portfolio, without using any information about how the single instruments contribute to the total risk. This leads to poor risk diversified strategies. In fact, it has been observed that these portfolios tend to be concentrated in a very limited subset of all the instruments. The criterion of selection of Definition 2.1.1 can lead to risky non-diversified portfolios in situations where the risk measure ρ is not coherent. We consider VaR in the case of Example 1.1.3 (from [17] Example 6.12)

Example 2.1.1. Consider portfolios $\mathbf{w} \in W_1^+$ of n defaultable bonds such as in Example 1.1.3. We assume additionally that the individual returns and default probabilities are all the same and equal to $r = 0.05$ and $p = 0.03$ respectively. The portfolio return is then given by

$$\mathbf{w}'\mathbf{R} = (1+r)\sum_{i=1}^{n} w_i I_i - 1, \tag{2.1.6}$$

and the expected portfolio return, which is given by $r(1-p) - p$, is independent of the chosen strategy $\mathbf{w}$. The mean-VaR optimal portfolio at level $\alpha = 0.05$ is therefore given by

$$\arg\min_{\mathbf{w}\in W_1^+} VaR_\alpha(\mathbf{w}'\mathbf{I}), \tag{2.1.7}$$

where $\mathbf{I} = (I_1, \ldots, I_n)$. Since for any strategy $\mathbf{w} \in W_1^+$ it holds $\mathbf{w}'\mathbf{I} \leq 1$, we have that $VaR_\alpha(\mathbf{w}'\mathbf{I}) \geq -1$. The lower bound can be achieved by investing all the money in one bond and taking no positions in the others. This strategy, where we set $w_1 = 1$ and $w_2 = \cdots = w_n = 0$, is therefore a mean-VaR optimal strategy, but our intuition about risk tells us that splitting our money in many bonds should be less risky. This happens because we are in a situation where Value-at-Risk is not subadditive, and so it does not represent the diversification effect. This is of course a simplistic situation but similar effects can be observed also in real-world settings. Better mean-ρ portfolios in this situation can be obtained by replacing VaR with a coherent risk measure such as ES.

Another well-known simple solution is the so called *minimum variance* portfolio, abbreviated MV portfolio,

$$\mathbf{w}^{mv} = \arg\min_{\mathbf{w}\in W} \mathbf{w}'\Omega\mathbf{w}, \tag{2.1.8}$$

where we minimize without imposing any condition on the overall portfolio return. This appears more robust than the mean-variance framework since we do not need to estimate the expected returns. Nevertheless, these portfolios still suffer from poor risk diversification. In the case W_l^+ for a certain leverage $l \geq 1$, a very heuristic approach is given by the so called *equally weighted portfolio*, abbreviated $\frac{1}{n}$*-portfolio*, where we simply set

$$w_i = \frac{l}{n}. \tag{2.1.9}$$

i.e we divide the capital lC equally among all the instruments. This portfolio appears balanced in the sense that the same weight is attributed to all assets, but in this way we do not take into account that the individual instruments may have significantly different individual risks. The central question we want answer is the following: "how do we build balanced portfolios that take into account the individual risks?". By the discussion of Section 1.3 we know that we can decompose the risk of any portfolio $\mathbf{w} \in W$ in (see Theorem 1.3.1)

$$\rho(\mathbf{w}) = \sum_{i=1}^{n} w_i \frac{\partial \rho}{\partial w_i}(\mathbf{w}) = \sum_{i=1}^{n} \rho_i(\mathbf{w}), \tag{2.1.10}$$

where the $\rho_i(\mathbf{w})$ are not just abstract terms that add up to the total risk, but they represent how the single assets contribute to the total portfolio risk. We want now to consider the portfolio risk components and "split the risk equally" rather than splitting the wealth equally. Thus, the ideal portfolio is defined as follows.

Definition 2.1.2. *An equally-weighted risk contributions portfolio, abbreviated ERC, is a strategy $\mathbf{w}^{erc} \in W$ such that*

$$\rho_i(\mathbf{w}^{erc}) = w_i^{erc} \frac{\partial \rho(\mathbf{w})}{\partial w_i}\bigg|_{\mathbf{w}=\mathbf{w}^{erc}} = w_j^{erc} \frac{\partial \rho(\mathbf{w})}{\partial w_j}\bigg|_{\mathbf{w}=\mathbf{w}^{erc}} = \rho_j(\mathbf{w}^{erc}) \tag{2.1.11}$$

for all $i, j \in \{1, \dots, n\}$.

The principle behind this is very intuitive: we equalize the risk contributions of the different portfolio components so that none of the assets will contribute more than the others to the overall portfolio risk. Dealing with risk contributions has become a standard practice for many investors, although there is not so much literature about ERC strategies. In the working article [13] the authors discuss the properties of ERC strategies using the standard deviation as risk measure. In particular they have shown that

ERC portfolios are located between the mean-variance strategies and the equally-weighted strategies in terms of volatility. There isn't however as far as we know any literature where other risk measures are considered. Provided that the risk measures satisfy the necessary assumptions there are no theoretical difficulties in applying the ERC principle on VaR, ES or other risk measures. Especially in the case of elliptical distributions, where by Theorem 1.4.3, the risk contributions can be computed explicitly and are very similar to the ones of the standard deviation. In the next section we consider ERC strategies from analytical and numerical point of view using the standard deviation risk measure. We will follow quite closely the work done in the paper [13]. We will then explain how the problem of finding ERC strategies looks like when other risk measures are considered.

2.2 ERC Strategies using Standard Deviation

2.2.1 The Problem

Assume that $Var[R_i] < \infty$ for all $i = 1, \dots, n$. Let Ω as in Section 1.2 denotes the positive definite covariance matrix of $\mathbf{R}$ and consider the standard deviation risk measure $\sigma(\mathbf{w}) = \sqrt{\mathbf{w}'\Omega\mathbf{w}}$, which is a positive homogenous partially differentiable risk measure defined on the whole Euclidean space $\mathbb{R}^n$. Let, as in the previous section, $C > 0$ and $W \subset \mathbf{R}^n$ be a fixed initial wealth and a fixed set of admissible strategies. We will often restrict ourselves to the case of W_l^+, mentioning explicitly when doing so. Recalling equation (1.4.3), the risk contributions for the standard deviation are given by

$$\rho_i(\mathbf{w}) = w_i \frac{\sum_{j\neq i} w_j \chi_{ij} \sigma_i \sigma_j + \sigma_i^2 w_i}{\sigma(\mathbf{w})}, \tag{2.2.1}$$

where we use the notation $\sigma_i^2 = Var[R_i]$ and $\chi_{ij} = \frac{Cov(R_i, R_j)}{\sigma_i \sigma_j}$. We look therefore for a portfolios $\mathbf{w}^{erc}$ such that

$$w_i^{erc} \sum_{k\neq i} w_k^{erc} \chi_{ik} \sigma_i \sigma_k + \sigma_i^2 w_i^{erc} = w_j^{erc} \sum_{k\neq j} w_k^{erc} \chi_{jk} \sigma_j \sigma_k + \sigma_j^2 w_j^{erc} \tag{2.2.2}$$

for all $i, j = 1, \dots, n$. Equation (2.2.2) can be written more compactly as $w_i^{erc}\big(\Omega\mathbf{w}^{erc}\big)_i = w_j^{erc}\big(\Omega\mathbf{w}^{erc}\big)_j$. If such a strategy exists is not immediately clear from the equations and the existence depends on the constraints. We will show that for the case of W_l^+ the ERC portfolio exists and is unique. If the ERC portfolio $\mathbf{w}^{erc} \in W$ exists, it means that there exists a constant $c > 0$ such that

$$\mathbf{w}^{erc} \otimes \Omega\mathbf{w}^{erc} = c\mathbf{1}, \tag{2.2.3}$$

where $\otimes$ denotes the element by element product. To develop some intuition about how the ERC strategy is selected we consider the following example.

Example 2.2.1. For the case of W_l^+, we rewrite equation (2.2.2) in a nicer way. Define,

$$\beta_i(\mathbf{w}) = \frac{1}{\sigma(\mathbf{w})} Cov\Big(R_i, \sum_{k=1}^{n} w_k R_k\Big) = \frac{\sum_{k\neq i} w_k \chi_{ik}\sigma_i\sigma_k + \sigma_i^2 w_i}{\sigma(\mathbf{w})}, \qquad (2.2.4)$$

and thus $\rho_i(\mathbf{w}) = w_i\beta_i(\mathbf{w})$. The condition $\rho_i(\mathbf{w}) = \rho_j(\mathbf{w})$ is thus equivalent to $w_i\beta_i(\mathbf{w}) = w_j\beta_j(\mathbf{w})$ for all i, j. Using the constraint that the portfolio weights add up to l, the ERC condition can then be equivalently reformulated in the following way.

$$\begin{aligned} &\sum_{j=1}^{n} w_j = l \Leftrightarrow \sum_{j=1}^{n} \beta_i(\mathbf{w}) w_j = \beta_i(\mathbf{w}) l \Leftrightarrow \sum_{j=1}^{n} \beta_i(\mathbf{w}) \frac{\beta_i(\mathbf{w}) w_i}{\beta_j(\mathbf{w})} = \beta_i(\mathbf{w}) l \\ &\Leftrightarrow w_i = \frac{\beta_i(\mathbf{w})^{-1} l}{\sum_{j=1}^{n} \beta_j(\mathbf{w})^{-1}}, \end{aligned} \qquad (2.2.5)$$

by definition of the ERC strategy we have $\frac{\rho_i(\mathbf{w})}{\sigma(\mathbf{w})} = \frac{1}{n}$, thus

$$\sum_{j=1}^{n} \beta_j(\mathbf{w})^{-1} = \sum_{j=1}^{n} \frac{w_j\sigma(\mathbf{w})}{\rho_j(\mathbf{w})} = n\sum_{j=1}^{n} w_j = nl, \qquad (2.2.6)$$

and therefore (2.2.2) is equivalent to the equations

$$\mathbf{w}_i = \frac{1}{n\beta_i(\mathbf{w})} \qquad (2.2.7)$$

for $i = 1, \ldots, n$. We can observe that the higher (lower) is the beta of the asset i, the lower (higher) is the proportion invested in i. This means that we dislike components with high volatility or high correlations with other assets. Note that the constraint $\sum_{i=1}^{n} w_i = l$ is only a normalizing one. Given any solution in $\mathbf{w} \in W^+$ we can find an ERC strategy in W_l^+ dividing the previous solution by $\frac{l}{\mathbf{1}'\mathbf{w}}$.

A general closed-form solution for the ERC strategy is not available, but in some very simple situations it is possible to compute the ERC portfolios explicitly.

Example 2.2.2. Consider a two asset case with constraints W_1^+. The covariance matrix is given by

$$\Omega = \begin{bmatrix} \sigma_1^2 & \chi\sigma_1\sigma_2 \\ \chi\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}, \tag{2.2.8}$$

where σ_1, σ_2 are the individual return standard deviations and χ denotes the correlation coefficient of R_1 and R_2. We look for a portfolio $\mathbf{w} = (w, 1-w)$ such that $\sigma_1(w, 1-w) = \sigma_2(w, 1-w)$. For the partial derivatives we have

$$\nabla_{\mathbf{w}}\sigma(w, 1-w) = \frac{(w\sigma_1^2 + (1-w)\chi\sigma_1\sigma_2, (1-w)\sigma_2^2 + w\chi\sigma_1\sigma_2)}{\sqrt{w^2\sigma_1^2 + 2\chi\sigma_1\sigma_2 w(1-w) + (1-w)^2\sigma_2^2}}, \tag{2.2.9}$$

and therefore the ERC portfolio must satisfy the following quadratic equation,

$$(\sigma_1^2 - \sigma_2^2)w^2 + 2\sigma_2^2 w - \sigma_2^2 = 0, \tag{2.2.10}$$

with solutions given by

$$w^{(1)} = \frac{\sigma_2}{\sigma_1 + \sigma_2}, \quad w^{(2)} = \frac{\sigma_2}{\sigma_2 - \sigma_1}. \tag{2.2.11}$$

the first solution is acceptable since both $w^{(1)}$ and $1 - w^{(1)}$ lie in the interval $[0, 1]$. For $w^{(2)}$ to be acceptable we need $\sigma_2 > \sigma_1$ but in this case for the other component of the portfolio we have,

$$1 - w^{(2)} = -\frac{\sigma_1}{\sigma_2 - \sigma_1} < 0, \tag{2.2.12}$$

and so the second solution of the quadratic equation does not lead to an ERC portfolio satisfying the desired constraints. Thus, we have a unique ERC portfolio for these constraints given by

$$\mathbf{w}^{erc} = \Big(\frac{\sigma_2}{\sigma_1 + \sigma_2}, \frac{\sigma_1}{\sigma_1 + \sigma_2}\Big). \tag{2.2.13}$$

Example 2.2.3. Let us consider a portfolio of n financial instruments for which we assume that

$$Cov(R_i, R_j) = \chi\sigma_i\sigma_j, \tag{2.2.14}$$

for all $i \neq j$, i.e. we assume that the returns have constant correlation coefficient. The risk contribution under this assumption for $i = 1, \ldots, n$ is then given by

$$\begin{aligned} \rho_i(\mathbf{w}) &= \frac{1}{\sigma(\mathbf{w})}\Big(w_i^2\sigma_i^2 + \chi\sum_{j\neq i} w_i w_j \sigma_i\sigma_j\Big) \\ &= \frac{w_i\sigma_i}{\sigma(\mathbf{w})}\Big((1-\chi)w_i\sigma_i + \chi\sum_{j=1}^{n} w_j\sigma_j\Big), \end{aligned} \tag{2.2.15}$$

where $\sigma_i^2 = Var[R_i]$. An ERC portfolio can therefore be obtained setting

$$w_i \sigma_i = w_j \sigma_j \tag{2.2.16}$$

for all i, j. Under the constraints W_l^+ , the above equation implies

$$l = \sum_{j=1}^{n} w_j = \sigma_i w_i \sum_{j=1}^{n} \sigma_j^{-1}, \tag{2.2.17}$$

and therefore we have the following explicit solution

$$w_i^{erc} = \frac{\sigma_i^{-1} l}{\sum_{j=1}^{n} \sigma_j^{-1}}. \tag{2.2.18}$$

2.2.2 Solutions by Numerical Optimization

The fact that no closed-form solution is available in general, and a solution does not need to exist for general constraints, motivates the formulation of the ERC portfolio selection problem as an *optimization problem.* In this way if the ERC strategy satisfying the constraints that equalizes all the single risks does not exist, the portfolio that equalizes the risks as best as possible (according to a specified criterion) will be chosen as *optimal strategy.* The MV portfolio we have seen in Section 2.1 is also chosen according to an optimality criterion, which is however completely different from the one we are going to introduce. We select the portfolio according to the following optimality criterion.

Definition 2.2.1. *Let $f : W \to [0, \infty[$ be a positive real valued function on W such that $f(\mathbf{w}) = 0$ only when $\mathbf{w}$ is an ERC portfolio. We define the ERC-optimal portfolio for f as*

$$\mathbf{w}^{opt} = \arg\min_{\mathbf{w} \in W} f(\mathbf{w}). \tag{2.2.19}$$

The function f represents a well-defined distance function relative to the ERC portfolio and we want to be close to the ERC portfolio in terms of f. Possible ways of choosing this function are for example the following.

$$
\begin{aligned}
f_1(\mathbf{w}) &= \sum_{i=1}^{n}\sum_{j=1}^{n}\big(w_i(\Omega\mathbf{w})_i - w_j(\Omega\mathbf{w})_j\big)^2 \\
&= 2n\sum_{i=1}^{n} w_i^2(\Omega\mathbf{w})_i^2 - 2\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j(\Omega\mathbf{w})_i(\Omega\mathbf{w})_j \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\big|\rho_i(\mathbf{w}) - \rho_j(\mathbf{w})\big|^2\sigma(\mathbf{w})^2, \\
f_2(\mathbf{w}) &= \sum_{i=1}^{n}\Big(\frac{\sigma(\mathbf{w})}{n} - \sigma_i(\mathbf{w})\Big)^2 = \frac{1}{\mathbf{w}'\Omega\mathbf{w}}\sum_{i=1}^{n}\Big(\frac{\mathbf{w}'\Omega\mathbf{w}}{n} - w_i(\Omega\mathbf{w})_i\Big)^2, \\
f_3(\mathbf{w}) &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\big|\sigma_i(\mathbf{w}) - \sigma_j(\mathbf{w})\big|.
\end{aligned}
\tag{2.2.20}
$$

The above optimization problem can be solved numerically using SQP (Sequential Quadratic Programming) algorithm. We briefly sketch how this algorithm works in Appendix C. We are going to use the Matlab Implementation of this method, which will be discussed later. We will use the optimality criteria f_1 or f_2 (which define L^2-distances), whereas f_3 (which defines an L^1-distance) is not applicable because the function is not everywhere differentiable. Often we will also be interested in fixing a certain risk target for the total portfolio risk and then finding the best possible strategy with total risk as near as possible to the target. This can be achieved for a certain risk target $\rho_0 > 0$ replacing f in Definition 2.2.1 by

$$
f_{\rho_0}(\mathbf{w}) = \sum_{i=1}^{n}\Big(\frac{\rho_0}{n} - \rho_i(\mathbf{w})\Big)^2, \tag{2.2.21}
$$

for which we have that $f(\mathbf{w}) = 0$ only if $\mathbf{w}$ is an ERC strategy with total portfolio risk equal to ρ_0. The formulation (2.2.19) of the portfolio selection is therefore flexible: we can consider other optimality conditions by modifying the function f. Let's consider the constraints W_l^+, and the corresponding optimization problem

$$
\mathbf{w}^{opt} = \arg\min f_1(\mathbf{w}) \quad \text{subject to} \quad \begin{cases} \mathbf{1}'\mathbf{w} = l \\ w_i \geq 0 \qquad i = 1,\ldots,n \end{cases} \tag{2.2.22}
$$

where f_1 is chosen such as in (2.2.20). We now prove the existence of an ERC strategy with these constraints switching to an alternative optimization problem.

Theorem 2.2.1. *For any $c \in \mathbb{R}$ the solution of the following optimization problem,*

$$\mathbf{w}^*(c) = \arg\min \sqrt{\mathbf{w}'\Omega\mathbf{w}} \quad \textit{subject to} \quad \begin{cases} \sum_{i=1}^n \ln(w_i) \geq c \\ w_i \geq 0 & i = 1, \ldots, n \end{cases} \tag{2.2.23}$$

is an ERC strategy. From this it follows that the ERC solution in the case of W_l^+ always exists.

Proof. We use a generalization of the Lagrange method that provides necessary conditions for optimization problems with inequality constraints such as the one above. These conditions are known under Kuhn-Tucker conditions and are presented formally in Appendix B. The Lagrangian function for this problem is given by

$$L(\mathbf{w}, \boldsymbol{\lambda}, \lambda_c) = \sqrt{\mathbf{w}\Omega\mathbf{w}} - \boldsymbol{\lambda}'\mathbf{w} - \lambda_c\Big(\sum_{i=1}^n \ln(w_i) - c\Big), \tag{2.2.24}$$

where the last two terms in the Lagrangian correspond to the inequalities $w_i \geq 0$ and $\sum_{i=1}^n \ln(w_i) \geq c$. Suppose that $\mathbf{w}^*$ is a solution of the problem for an arbitrary constant c. By Appendix B Theorem B.1 there are Lagrange multipliers λ^* and λ_c^* that satisfy the first order condition

$$\nabla_{\mathbf{w}} L(\mathbf{w}^*, \boldsymbol{\lambda}^*, \lambda_c^*) = \frac{\Omega\mathbf{w}^*}{\sqrt{\mathbf{w}^{*\prime}\Omega\mathbf{w}^*}} - \boldsymbol{\lambda}^* - \lambda_c^*\Big(\frac{1}{w_1^*}, \ldots, \frac{1}{w_n^*}\Big) = \mathbf{0}, \tag{2.2.25}$$

and the Kuhn-Tucker conditions

$$\begin{cases} w_i^* \geq 0 \;\; \text{for all} \;\; i = 1, \ldots, n \\ \sum_{i=1}^n \ln(w_i^*) - c \geq 0 \\ \lambda_i^*, \lambda_c^* \geq 0 \;\; \text{for all} \;\; i = 1, \ldots, n \\ \lambda_i^* w_i^* = 0 \;\; \text{for all} \;\; i = 1, \ldots, n \\ \lambda_c^*\Big(\sum_{i=1}^n \ln(w_i^*) - c\Big) = 0. \end{cases} \tag{2.2.26}$$

Since $\ln(w_i^*)$ is not defined for $w_i^* = 0$ we have $w_i^* > 0$, and so it follows from the above inequalities that $\lambda_i^* = 0$ for all $i = 1, \ldots, n$. Note that if $\lambda_c^* = 0$ then by (2.2.25) it follows that $\Omega\mathbf{w}^* = 0$, hence $\mathbf{w}^* = 0$, which cannot be. Thus, $\lambda_c^* > 0$, and therefore the constrained $\sum_{i=1}^n \ln(\mathbf{w}^*) = c$ must be necessarily reached. The solution satisfies, see (2.2.25),

$$w_i^* \frac{\partial \sigma}{\partial w_i}(\mathbf{w}^*) = \lambda_c^*, \tag{2.2.27}$$

with $\sigma(\mathbf{w}^*)$ denoting as before the standard deviation, for all $i = 1 \ldots, n$. This proves that the solution of (2.2.23) has equal contributions to risk.

Moreover this optimization problem belongs to the class of well-known problem of minimizing a quadratic function subject to lower convex bound which has solution $\mathbf{w}^* \in W^+$. To get the solution in W_l^+ we just need to normalize the obtained solution in the following way,

$$\frac{l}{\sum_{i=1}^n w_i^*}\mathbf{w}^*. \tag{2.2.28}$$

□

To solve the problem numerically the optimization problem (2.2.22) is better since the inequality constraints are given by linear functions. The authors of [13] suggest in order to obtain better numerical results (i.e. results with smaller $f(\mathbf{w}^*)$) to replace the equality condition in (2.2.22) with the less restrictive condition $\mathbf{1}'\mathbf{w} \geq k$ for an arbitrary constant $k > 0$, and then normalize the solution of the modified problem such as in (2.2.28). On the other side the optimization problem (2.2.23) allows us to show that there is a unique ERC solution in W_l^+ (provided that Ω is a positive definite covariance matrix). When we allow for short selling, relaxing the condition that $w_i \geq 0$, other ERC solutions may be found. The only thing that remains to do in order to apply this principle is estimating the covariance matrix Ω. We are going to consider this issue in Chapter 3.

2.2.3 ERC vs. MV and $\frac{1}{n}$ Strategies

These three strategies lead in general to different portfolio weights and the aim of this section is compare them, in particular we are going to expose the main theoretical result of [13] which states that the ERC strategies are located in the middle between MV and $\frac{1}{n}$ strategies in terms of volatilities and the ERC principle appears therefore as a good substitute for the classical selection principles. Let us go back to the simple two assets case and compare the portfolio weights of the three different strategies.

Example 2.2.4. Consider the two asset case of Example 2.2.1 where there's unique ECR solution is given by

$$\mathbf{w}^{erc} = \Big(\frac{\sigma_2}{\sigma_1+\sigma_2}, \frac{\sigma_1}{\sigma_1+\sigma_2}\Big). \tag{2.2.29}$$

The portfolio variance is then given by

$$\begin{aligned}\sigma(w,1-w)^2 &= (w,1-w)\begin{bmatrix}\sigma_1^2 & \chi\sigma_1\sigma_2\\ \chi\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix}\begin{pmatrix}w\\ 1-w\end{pmatrix}\\ &= \sigma_1^2 w^2 + 2\chi\sigma_1\sigma_2 w(1-w) + \sigma_2^2(1-w)^2,\end{aligned} \tag{2.2.30}$$

with first order derivative with respect to w is given by

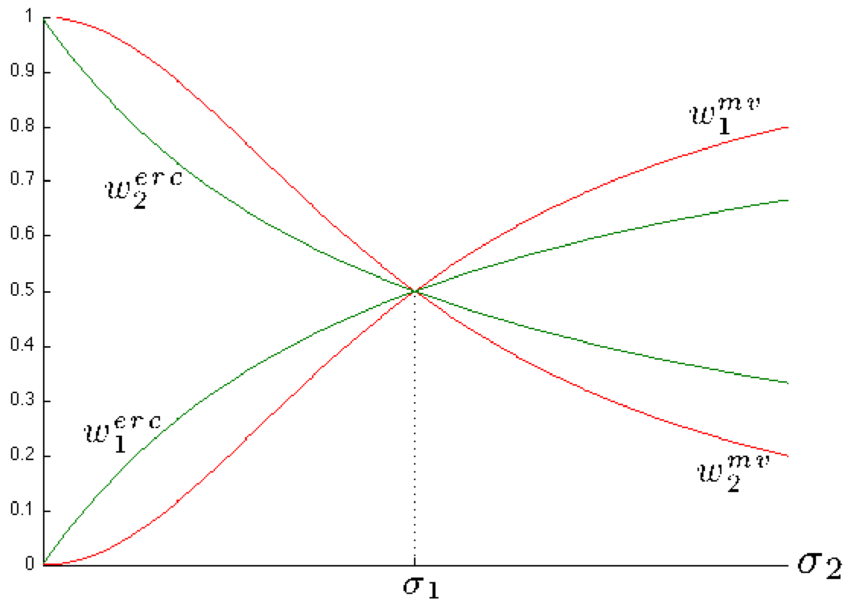


Figure 2.1: Portfolio weights as a function of σ_2.

$$\frac{d}{dw}\sigma(w,1-w)^2 = 2w(\sigma_1^2 - 2\chi\sigma_1\sigma_2 + \sigma_2^2) + 2\chi\sigma_1\sigma_2 - 2\sigma_2^2. \tag{2.2.31}$$

For the second derivative we have since $\chi \leq 1$,

$$\begin{aligned} \frac{d^2}{dw^2}\sigma(w,1-w)^2 &= 2(\sigma_1^2 - 2\chi\sigma_1\sigma_2 + \sigma_2^2) \\ &\geq 2(\sigma_1^2 - 2\sigma_1\sigma_2 + \sigma_2^2) = 2(\sigma_1 - \sigma_2)^2 \geq 0, \end{aligned} \tag{2.2.32}$$

so the minimum variance portfolio is given by

$$\mathbf{w}^{mv} = \frac{1}{\sigma_1^2 - 2\chi\sigma_1\sigma_2 + \sigma_2^2}\big(\sigma_2^2 - \chi\sigma_1\sigma_2, \sigma_1^2 - \chi\sigma_1\sigma_2\big). \tag{2.2.33}$$

For the $\frac{1}{n}$ strategy we have $\mathbf{w}^{\frac{1}{n}} = (\frac{1}{2}, \frac{1}{2})$. It follows immediately from the equation (2.2.29) that $\mathbf{w}^{erc} = \mathbf{w}^{\frac{1}{n}}$ if $\sigma_1 = \sigma_2$. Using equation (2.2.33) it is also evident that $\mathbf{w}^{mv} = \mathbf{w}^{\frac{1}{n}}$ if $\sigma_1 = \sigma_2$. Some simple algebra using equations (2.2.33) also shows that $\mathbf{w}^{mv} = \mathbf{w}^{erc}$ in the case $\sigma_1 = \sigma_2$ or $\chi = -1$.
The three strategies coincide in the case that both instruments have the same volatilities. When the individual risks differ the weights of the three strategies differ as well, since we deal with this risk difference between the two assets differently: in the $\frac{1}{n}$ strategy we do not care and divide our wealth equally, in the MV approach we minimize the total portfolio risk and

with the ECR principle we equalize the individual risk components, see also Figure 1.

When considering more assets the problem of comparing the results becomes more involved. Some results are however available. It is for example possible to show that the ERC portfolio coincides with the MV portfolio when the correlation coefficient is minimal. Precisely we have the following ([13] A.1).

Theorem 2.2.2. *Assume that the covariance structure of* $\mathbf{R}$ *has constant correlation coefficient* χ*. If the lower bound* $\chi = -\frac{1}{n-1}$ *for the correlation coefficient is achieved then* $\mathbf{w}^{mv} = \mathbf{w}^{erc}$*.*

Proof. By assumption the covariance matrix has the following form,

$$\Omega = \Sigma \otimes (\chi \mathbf{1}\mathbf{1}' + (1-\chi) I_n), \tag{2.2.34}$$

where $\otimes$ denotes the element by element product, $\Sigma_{ij} = \sigma_i \sigma_j$ and χ is a constant. For the inverse it holds

$$\Omega^{-1} = \bar{\Sigma} \otimes \frac{\chi \mathbf{1}\mathbf{1}' - \big((n-1)\chi + 1\big) I_n}{(n-1)\chi^2 - (n-2)\chi - 1}, \tag{2.2.35}$$

where $\bar{\Sigma}_{ij} = \frac{1}{\sigma_i \sigma_j}$. This follows because,

$$\begin{aligned}
&(\chi \mathbf{1}\mathbf{1}' + (1-\chi) I_n) \frac{\chi \mathbf{1}\mathbf{1}' - \big((n-1)\chi + 1\big) I_n}{(n-1)\chi^2 - (n-2)\chi - 1} \\
&= \frac{\chi^2 n \mathbf{1}\mathbf{1}' - \big((n-1)\chi + 1\big)\chi \mathbf{1}\mathbf{1}' + (1-\chi)\chi \mathbf{1}\mathbf{1}' - (1-\chi)\big((n-1)\chi + 1\big) I_n}{(n-1)\chi^2 - (n-2)\chi - 1} \\
&= I_n.
\end{aligned} \tag{2.2.36}$$

The MV solution is given by

$$w_i^{mv} = \frac{(\Omega^{-1}\mathbf{1})_i}{\mathbf{1}'\Omega^{-1}\mathbf{1}} = \frac{-\big((n-1)\chi + 1\big)\sigma_i^{-2} + \chi \sum_{j=1}^n (\sigma_i \sigma_j)^{-1}}{\sum_{k=1}^n \Big(- \big((n-1)\chi + 1\big)\sigma_k^{-2} + \chi \sum_{j=1}^n (\sigma_k \sigma_j)^{-1} \Big)}. \tag{2.2.37}$$

Substituting $\chi = -\frac{1}{n-1}$ we find

$$w_i^{mv} = \frac{\sum_{j=1}^n (\sigma_i \sigma_j)^{-1}}{\sum_{k=1}^n \sum_{j=1}^n (\sigma_k \sigma_j)^{-1}} = \frac{\sigma_i^{-1}}{\sum_{k=1}^n \sigma_k^{-1}} = w_i^{erc}. \tag{2.2.38}$$

□

The following which is the main theorem tells us that the ERC portfolio may be viewed as a compromise between the minimum variance and the equally weighted portfolio.

Theorem 2.2.3. *For W_l^+ being the set of admissible strategies and a real constant c consider the following optimization problem,*

$$\mathbf{w}^{opt}(c) = \arg\min \sqrt{\mathbf{w}'\Omega\mathbf{w}} \quad \textit{subject to} \quad \begin{cases} \sum_{i=1}^{n} \ln(w_i) \geq c \\ \mathbf{1}'\mathbf{w} = l \\ w_i \geq 0 \quad i = 1, \ldots, n \end{cases} . \tag{2.2.39}$$

There exists a certain $c^{erc} \in [-\infty, -n \ln n]$ such that

$$\mathbf{w}^{opt}(c^{erc}) = \mathbf{w}^{erc}. \tag{2.2.40}$$

In particular, this implies that the volatilities of the $\frac{1}{n}$, MC and ERC strategies are ordered as follows,

$$\sigma(\mathbf{w}^{mv}) \leq \sigma(\mathbf{w}^{erc}) \leq \sigma(\mathbf{w}^{\frac{1}{n}}). \tag{2.2.41}$$

Proof. First observe that if we set $c = -\infty$ there is no additional inequality constraint and the above optimization problem is equivalent to the MV problem. Thus we have $\mathbf{w}^{opt}(-\infty) = \mathbf{w}^{mv}$. Secondly, applying Jensen's Inequality we find

$$\frac{1}{n}\sum_{i=1}^{n} \ln(w_i) \leq \ln\Big(\frac{1}{n}\sum_{i=1}^{n} w_i\Big) = \ln(l) - \ln(n), \tag{2.2.42}$$

and so we have the bound $\sum_{i=1}^{n} \ln(w_i) \leq n\big(\ln(l) - \ln(n)\big)$ for the additional constraint. This bound is achieved for $w_i = \frac{l}{n}$ and therefore

$$\mathbf{w}^{opt}\Big(n\big(\ln(l) - \ln(n)\big)\Big) = \mathbf{w}^{\frac{1}{n}}. \tag{2.2.43}$$

Observe also that for $c_1 \leq c_2$ it holds $\sigma\big(\mathbf{w}^{opt}(c_1)\big) \leq \sigma\big(\mathbf{w}^{opt}(c_2)\big)$. This is because the additional constraint becomes more restrictive as the constant c grows. If we now show that there is a constant $c^{erc} \in [-\infty, n\big(\ln(l) - \ln(n)\big)]$ corresponding to the ERC strategy, then the inequality (2.2.29) follows directly form the above observations. The existence of such a constant follows directly from the work done in the proof of Theorem 2.2.1. For an arbitrary c the solution $\mathbf{w}^*$ of (2.2.22) is an ERC strategy. The ERC solution in W_l^+ obtained normalizing $\mathbf{w}^*$ as

$$\mathbf{w}^{erc} = \frac{l}{\sum_{i=1}^{n} w_i^*}\mathbf{w}^*, \tag{2.2.44}$$

is then the solution of (2.2.39) for $c^{erc} = c + n\ln(l) - n\ln\Big(\sum_{i=1}^{n} w_i^*\Big)$. By Jensen's inequality,

$$\ln\Big(\sum_{i=1}^{n} w_i^*\Big) \geq \frac{c}{n} + \ln(n), \tag{2.2.45}$$

where we used that $\sum_{i=1}^{n} \ln(w_i^*) = c$, as seen in the proof of Theorem 2.2.1. The constant c^{erc} satisfies then

$$c^{erc} \leq n\big(\ln(l) - \ln(n)\big). \tag{2.2.46}$$

This completes the proof. □

It is interesting to observe that the ERC portfolio can be obtained from (2.2.39) minimizing the standard deviation as in the MV approach under the additional condition

$$\sum_{i=1}^{n} \ln(w_i) \geq c^{erc}, \tag{2.2.47}$$

that can be interpreted as some kind of diversification requirement.

2.3 Other Risk Measures

Obviously the approach of Definition 2.2.1 can be applied to find the optimal portfolio numerically also when using other risk measures instead of standard deviation. It suffices to replace $\rho_i(\mathbf{w}) = w_i\big(\Omega\mathbf{w}\big)_i$ in the above equations with the corresponding risk contribution obtained by differentiating the particular risk measure under consideration

$$RC_i^{\rho}(\mathbf{w}) = w_i \frac{\partial \rho}{\partial w_i}(\mathbf{w}). \tag{2.3.1}$$

We discuss the formulation of the optimal portfolio problem in the case of elliptical distributed returns and in the case of Value-at-Risk and Expected Shortfall.

2.3.1 Elliptical Models

If we assume that $\mathbf{R} \sim E_n(\boldsymbol{\mu}, \Omega, \psi)$ there are no particular problems in generalizing the above discussion to other risk measures. Fix a set of acceptable strategies W and a certain risk measure ρ. We assume in this section that ρ is a positive homogeneous, translation invariant and partially differentiable risk measure. We also assume that ρ depends only on the distribution of the profit. Then, by Theorem 1.4.3, the risk contributions for $i = 1, \ldots, n$ are given by

$$RC_i^{\rho}(\mathbf{w}) = \rho_i(\mathbf{w})q - \mu_i, \tag{2.3.2}$$

where $\rho_i(\mathbf{w})$ is the risk contribution of the standard deviation and $q_{\rho,\psi} = q = \rho(Y_1)$ is a constant that depends on the particular ψ of the elliptical model and on the risk measure under consideration. For example: for the multivariate normal model and VaR_α we have $q = -\Phi^\alpha(\alpha)$ (see Example 1.4.1). The ERC portfolio problem in this case consists in finding a strategy $\mathbf{w}^{erc}$ such that for all $i, j = 1, \ldots, n$

$$\rho_i(\mathbf{w}^{erc}) - \rho_j(\mathbf{w}^{erc}) = -\frac{1}{q}(\mu_i - \mu_j). \tag{2.3.3}$$

This representation provides some intuition about how the ERC strategy is chosen in the elliptical case. We can observe that here information about the asset returns is used to find the ERC portfolio. The portfolio is chosen so that for every pair of assets the ratio between the difference of the individual standard deviation risk contributions and the difference of the returns is constant. From this it is also evident that the ERC portfolios using ρ and σ as risk measures coincide only if all the assets returns are equal. The ERC problem for σ is therefore a special case of the one for ρ. Also in this case no closed-form analytical solutions are available. We can formulate the optimal portfolio problem in this case as in Definition 2.2.1 using analogous the functions f_1, f_2 or f_{ρ_0} where ρ_i is replaced by RC_i^ρ. The only things that remains to do is estimating the covariance matrix Ω and the vector of expected returns $\boldsymbol{\mu}$. We are going to consider an algorithm to estimate these parameters in the elliptical model based on historical data in Chapter 3.

2.3.2 VaR and ES for General Models

Assuming that the distribution of $\mathbf{R}$ satisfies the necessary assumptions we have by the work done in Section 1.4 the following risk contributions for a portfolio $\mathbf{w}$,

$$\begin{aligned} RC_i^{VaR_\alpha}(\mathbf{w}) &= -w_i E\Big[R_i \Big| \mathbf{w}'\mathbf{R} = -VaR_\alpha(\mathbf{w})\Big], \\ RC_i^{ES_\alpha}(\mathbf{w}) &= -w_i \frac{1}{\alpha}\Big\{E\big[R_i 1_{\{\mathbf{w}'\mathbf{R} \le q_\alpha(\mathbf{w}'\mathbf{R})\}}\big] \\ &\quad + E\big[R_i | \mathbf{w}'\mathbf{R} = q_\alpha(\mathbf{w}'\mathbf{R})\big]\big(\alpha - P[\mathbf{w}'\mathbf{R} \le q_\alpha(\mathbf{w}'\mathbf{R})]\big)\Big\}, \end{aligned} \tag{2.3.4}$$

where the last expression in case of continuous profit distribution simplifies to

$$RC_i^{ES_\alpha}(\mathbf{w}) = -w_i E\Big[R_i \Big| \mathbf{w}'\mathbf{R} \le -VaR_\alpha(\mathbf{w})\Big]. \tag{2.3.5}$$

The main difficulty consists therefore in finding estimators for the above expressions. Once such estimators are available we can find the optimal strategy as explained in Section 2.2. In Chapter 3 we are going to treat the problem of estimating (2.3.4) and (2.3.5). This quantities can be estimated with historical data or using Monte Carlo Methods. There is some new literature proposing methods to estimate these risk contributions by simulation, see for example [20]. We are going to discuss some of the methods based on historical data in the next chapter.

Chapter 3

Estimating Risk Contributions

3.1 Covariance Matrix Estimation by Historical Data

In this section we present some covariance matrix estimators for asset returns. Suppose we have the return observations $\mathbf{r}_1, \ldots, \mathbf{r}_N$, where, for $i = 1, \ldots, N$ and $j = 1, \ldots, n$, r_{ij} denotes the observation i of the instrument j. We assume that these returns are independent observations (for different i) of the same distribution, but no particular form of the distribution such as normal is assumed. We also assume that the distribution is such that second moments exist. The independence assumption of asset returns through time is actually not verified but it is an acceptable approximation. The question of constructing covariance matrix estimators using the data is a very important issue in finance, especially in the context of this thesis where in order to estimate the risk contributions of the assets in the portfolio we often need to estimate the covariances between the asset returns. The standard estimators for the expected returns and the covariances are the *sample estimators*, defined by

$$\hat{\mathbf{r}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{r}_i, \qquad \hat{\Omega} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{r}_i - \hat{\mathbf{r}})(\mathbf{r}_j - \hat{\mathbf{r}})'. \tag{3.1.1}$$

The estimator $\hat{\mathbf{r}}$ is an unbiased estimator for the vector of expected returns, but $\hat{\Omega}$ is not unbiased for Ω. An unbiased version can be obtained by $\frac{n}{n-1}\hat{\Omega}$ as it can immediately be verified computing $E[\hat{\Omega}]$. The sample covariance matrix estimator $\hat{\Omega}$ has the appealing property of being the maximum likelihood estimator under normally distributed asset returns, i.e. the sample covariances are the most likely parameters value given the data. Covariance matrices are always positive semi-definte. The sample covariance matrix estimator produces always a positive semi-definite matrix as shown in the following example.

Example 3.1.1. Let $\mathbf{r}_1, \ldots, \mathbf{r}_N \in \mathbb{R}^n$ be N data vectors. Define a random vector $\mathbf{X}$ with $X_j \in \{r_{1j}, \ldots, r_{Nj}\}$ for $j = 1, \ldots, n$ and

$$P[\mathbf{X}_i = \mathbf{r}_i] = \frac{1}{N}, \tag{3.1.2}$$

for $i = 1, \ldots, N$. Note that, if some of the data vectors are equal to each others, the above probabilities need to be modified. The random vector $\mathbf{X}$ has the following properties

$$\begin{aligned} E[\mathbf{X}] &= \frac{1}{N}\sum_{i=1}^{N} \mathbf{r}_i = \hat{\mathbf{r}}, \\ Cov(\mathbf{X}) &= E\big[(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])'\big] = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{r}_i - \hat{\mathbf{r}})(\mathbf{r}_i - \hat{\mathbf{r}})' = \hat{\Omega}. \end{aligned} \tag{3.1.3}$$

The true covariance matrix of this random vector is equal to the sample covariance matrix of the data set $\mathbf{r}_1, \ldots, \mathbf{r}_N$. Since a true covariance matrix is known to be positive semi-definite, then it follows that also the sample covariance matrix estimator is also positive semi-definite.

Despite these nice properties sample covariance matrices present some serious deficiencies. The sample estimator let the data speak and only the data. But what about if we are in a situation where there are not enough data to give full weight to the them? In the situation where we have more assets than data the sample estimation is obviously problematic since there are more parameters to estimate than historical returns. The sample covariance matrix can be written in matrix notation as

$$\hat{\Omega} = \frac{1}{N} R'\Big(I - \frac{1}{N}\mathbf{1}\mathbf{1}'\Big)R, \tag{3.1.4}$$

where $R = (r_{ij})_{1 \le i \le N, 1 \le j \le n}$ is the matrix of observations, $I \in \mathbb{R}^{N \times N}$ is the identity matrix and $\mathbf{1} \in \mathbb{R}^N$ is the vector of ones. From the above equation it follows that whenever n exceeds $N - 1$, the sample covariance matrix is rank-deficient. This is because the rank of $\hat{\Omega}$ is at most equal to the rank of the matrix $I - \frac{1}{N}\mathbf{1}\mathbf{1}'$, which is $N - 1$. Whenever we have $n \geq N$ the sample estimator always lead to a singular matrix even if the true covariance matrix is assumed to be non-singular. Because of this, we need to modify the sample estimator. Small samples situations are often considered in the context of portfolio construction where we want for example to choose portfolio weights among several hundreds of financial instruments using the daily observations of the last three months. Obtaining a singular estimated covariance matrix is very problematic in the context of this thesis. For example, it is problematic in the context of risk contributions for the standard deviation risk measure where we assume that $\sigma^2(\mathbf{w}) = \mathbf{w}'\Omega\mathbf{w} > 0$, i.e. that the covariance matrix is positive definite. The method we want to present is usually referred to in the literature as *shrinkage* and it consists in combining the asymptotically unbiased sample covariance matrix estimator $\hat{\Omega}$ with another biased highly structured estimator, denoted by F. We want for our final estimator to be a compromise between the two estimators and we consider therefore the credibility weighted average

$$\Omega^{(\alpha)} = \alpha F + (1-\alpha)\hat{\Omega}, \tag{3.1.5}$$

for $\alpha \in (0, 1)$. The estimator F is referred to in the literature as the *prior estimator* or *shrinking target* and α as *shrinkage constant* or *credibility weight*, since the sample estimator is "shrunk" towards the structured estimator. It is important to explain more precisely what we mean when say that we want F to be an estimator with lot of structure: we want F to be a matrix that involve only a small number of free parameters and that at the same time captures important information on the true covariance matrix. This approach is very similar to the one used in Bayesian statistics where the same convex linear combinations are considered to give at the same time weight to the data and to other beliefs about the problem under consideration. Choosing a proper shrinking target is an issue that depends on the estimation problem under consideration, in the context of estimating stock returns covariances Ledoit and Wolf [11] have for example proposed a prior based on Sharpe's single-index model for stocks returns. However, the methodology proposed by the authors does not rely strongly on the chosen target and can be applied more generally. Ledoit and Wolf [12] present in a sequent paper an estimator that is quite easy to implement and which is based on a prior covariance matrix with constant correlation coefficient. We are going to formulate and use this estimator which is known to perform well in the context of portfolio construction. Another advantage of this shrinkage estimator is that it is always positive definite, since it is a convex combination of an estimator that is positive definite (the shrinkage target F) and an estimator that is positive semi-definite (the sample estimator). We start therefore with the *constant correlation model* as prior which states that all the pairwise asset correlations are identical. This clearly imposes a particular covariance matrix structure that can be estimated by using the average of all the sample correlations as common constant correlation together with the sample variances. Formally, if we denote by χ_{ij} the true correlation coefficient between assets i and j, then we define the true constant correlation matrix by

$$\phi_{ii} = \Omega_{ii}, \quad \phi_{ij} = \chi\sqrt{\Omega_{ii}\Omega_{jj}}, \tag{3.1.6}$$

where Ω denotes the true covariance matrix and

$$\chi = \frac{2}{(n-1)n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\chi_{ij}. \tag{3.1.7}$$

Our shrinking target is then defined as the sample constant correlation matrix

$$F_{ii} = \hat{\Omega}_{ii}, \quad F_{ij} = \hat{\chi}\sqrt{\hat{\Omega}_{ii}\hat{\Omega}_{jj}}, \tag{3.1.8}$$

where

$$\hat{\chi} = \frac{2}{(n-1)n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\hat{\chi}_{ij}, \quad \hat{\chi}_{ij} = \frac{\hat{\Omega}_{ij}}{\sqrt{\hat{\Omega}_{ii}\hat{\Omega}_{jj}}}. \tag{3.1.9}$$

The obvious question is now how the parameter α should be selected. We use the following optimality criterion.

Definition 3.1.1. *The optimal shrinkage intensity α^* is defined as*

$$\alpha^* = \arg\min_{\alpha\in[0,1]} E\Big[\big|\Omega - \Omega^{(\alpha)}\big|_F^2\Big], \tag{3.1.10}$$

where $\Omega^{(\alpha)}$ defines, as a function of the observations, a random variable and $|\cdot|_F$ denotes the Frobenius norm defined by

$$|A|_F^2 = \sum_{i,j=1}^{n} A_{ij}^2, \tag{3.1.11}$$

for any matrix $A \in \mathbb{R}^{n\times n}$.

The key result, which provides an asymptotic description of the optimal shrinkage intensity, and is independent of the particular structured estimator under consideration is [11] Theorem 1. This result states (under some assumptions about finite variances in order to be able to apply the central limit theorem) that, for the following quantities,

$$\begin{aligned}
\pi &= \sum_{i,j=1}^{n} AsyVar[\sqrt{N}\hat{\Omega}_{ij}], \\
\rho &= \sum_{i,j=1}^{n} AsyCov(\sqrt{N}F_{ij}, \sqrt{N}\hat{\Omega}_{ij}], \\
\gamma &= \sum_{i,j=1}^{n} \big(\phi_{ij} - \Omega_{ij}\big)^2,
\end{aligned} \tag{3.1.12}$$

we have,

$$\alpha^* = \frac{1}{N}\frac{\pi - \rho}{\gamma} + O\Big(\frac{1}{N^2}\Big), \tag{3.1.13}$$

as $N \to \infty$. This result allows to estimate the optimal shrinkage intensity by finding consistent estimators $\hat{\pi}$, $\hat{\rho}$ and $\hat{\gamma}$ for the above quantities. Consistent estimators for π and γ can be found independently of the chosen prior and are given by

$$\begin{aligned} \hat{\pi} &= \sum_{i,j=1}^{n} \hat{\pi}_{ij}, \quad \hat{\pi}_{ij} = \frac{1}{N}\sum_{k=1}^{T}\Big[\Big(r_{ki} - \frac{1}{N}\sum_{l=1}^{N} r_{li}\Big)\Big(r_{kj} - \frac{1}{N}\sum_{l=1}^{N} r_{lj}\Big) - \hat{\Omega}_{ij}\Big]^2, \\ \hat{\gamma} &= \sum_{i,j=1}^{N}\big(F_{ij} - \hat{\Omega}_{ij}\big)^2, \end{aligned} \tag{3.1.14}$$

as proven in [11] Lemma 1 and Lemma 3. Expressions to estimate ρ need however to be readjusted for different priors. A consistent estimator in this case is computed in [12] Appendix B and given by,

$$\hat{\rho} = \sum_{i=1}^{n} \hat{\pi}_{ii} + \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} \frac{\hat{\chi}}{2}\left(\sqrt{\frac{\hat{\Omega}_{jj}}{\hat{\Omega}_{ii}}}\hat{\vartheta}_{ii,ij} + \sqrt{\frac{\hat{\Omega}_{ii}}{\hat{\Omega}_{jj}}}\hat{\vartheta}_{jj,ij}\right), \tag{3.1.15}$$

where $\vartheta_{ii,ij}$ and $\vartheta_{jj,ij}$ are the consistent estimator of the asymptotic covariances $AsyCov(\sqrt{N}\hat{\Omega}_{ii}, \sqrt{N}\hat{\Omega}_{ij})$ and $AsyCov(\sqrt{N}\hat{\Omega}_{jj}, \sqrt{T}\hat{\Omega}_{ij})$ given in [12] Appendix B. Now we have all the elements to apply this shrinkage estimator. We are going to use this estimator when estimating risk contributions that require covariance matrix estimation.

3.2 Estimating VaR and ES Risk Contributions by Historical Data

As in the previous section we use the notation r_{ij}, for $i = 1, \dots, N$ and $j = 1, \dots, n$, to indicate independent (w.r.t. i) return observations of the same distribution. We also assume that the distribution of the returns is nice enough so that the representations (2.3.4) and (2.3.5) for the risk contributions are available. We want to use this conditional expectations to estimate the risk contributions using historical return data. For certain portfolio $\mathbf{w} \in \mathbb{R}^n$, we define the vector of observed portfolio returns $\mathbf{x} = (x_1, \dots, x_N)$ by

$$x_i = \mathbf{w}'\mathbf{r}_i, \tag{3.2.1}$$

for $i = 1, \dots, N$. If the number of observations N is large, then by the law of large numbers the portfolio return distribution can be approximated by the empirical distribution based on the historical portfolio return data:

$$P(X \leq x) \approx \frac{\#\{i = 1, \dots, N | x_i \leq x\}}{N}. \tag{3.2.2}$$

The VaR and ES risk measures can then be approximated with

$$\begin{aligned} VaR_\alpha(\mathbf{w}) &\approx -x^{sort}_{\lceil \alpha N \rceil}, \\ ES_\alpha(\mathbf{w}) &\approx -\frac{1}{\lfloor \alpha N \rfloor} \sum_{i=1}^{\lfloor \alpha N \rfloor} x^{sort}_i, \end{aligned} \tag{3.2.3}$$

where $x^{sort}_1 \leq \cdots \leq x^{sort}_N$ denote the sorted portfolio returns. From (2.3.5) it is clear how to use historical data to approximate ES risk contribution.

$$RC^{ES_\alpha}_i(\mathbf{w}) = -E[R_i | X \leq -VaR_\alpha(X)] \approx -\frac{1}{\lfloor \alpha N \rfloor} \sum_{k=1}^{\lfloor \alpha N \rfloor} r^{sort}_{ki}, \tag{3.2.4}$$

where r^{sort}_{ki} are the corresponding asset returns to the sorted portfolio returns. The estimation of VaR is because of the equality condition much more problematic. A possible way of doing it, it is using a gaussian kernel estimation for some fixed bandwidth $b > 0$:

$$RC^{VaR_\alpha}_i(\mathbf{w}) = -E[R_i | X = -VaR_\alpha(X)] \approx -\frac{\sum_{k=1}^N r_{ki} \varphi\left(\frac{-VaR_\alpha(X) - x_k}{b}\right)}{\sum_{k=1}^N \varphi\left(\frac{-VaR_\alpha(X) - x_k}{b}\right)}, \tag{3.2.5}$$

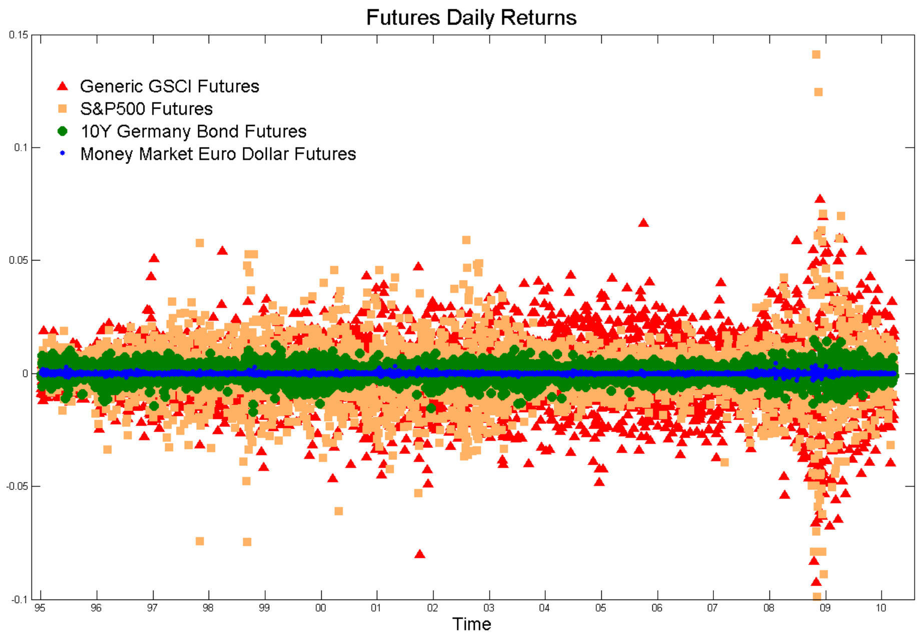
where $VaR_\alpha(X)$ can be approximated as in (3.2.3). The main problem of this method is that there is no general rule for fixing b and the choice of the bandwidth is quite crucial in order to get accurate approximation. We will fix the parameter b on a case-by-case basis by numerical experiments. We can measure how accurate the above approximation is by comparing $\sum_{i=1}^n RC^{VaR_\alpha}_i(\mathbf{w})$ with the value of $VaR_\alpha(\mathbf{w})$ obtained with (3.2.3).

Chapter 4

Illustrations in Practice

4.1 A first example: four asset classes

As a first illustration we consider futures portfolios consisting of S&P500 Futures, 10Y Germany Bond Futures, Money Market Euro Dollar Futures and Generic GSCI Futures. We consider an observation period of 15 years using daily observations from 02.01.1995 to 03.03.2010, for a total of $N = 3958$ observations. We denote the observed daily returns as we did in Chapter 3 by $\mathbf{r}_1, \ldots, \mathbf{r}_N$, with $\mathbf{r}_i = (r_{i1}, \ldots, r_{in})$, where in this example $n = 4$. The statistics of the daily returns of the 4 assets over this period are the following.



	S&P500	Bond	MM	GSCI
Mean	0.022%	0.016%	0.003%	0.016%
Ann. Mean	5.58%	4.14%	0.69%	4.21%
Max.	14.11%	1.48%	0.47%	7.69%
Min.	-9.88%	-1.71%	-0.32%	-9.24%
Standard deviation	1.28%	0.33%	0.04%	1.46%
Ann. standard deviation	20.26%	5.17%	0.64%	23.14%
Sharpe	0.28	0.80	1.08	0.18
Skewness	18.36%	-25.53%	1.0397	-14.58%
Kurtosis	14.229	4.5827	20.7445	5.7404

We also compute the sample correlations using the data over this period. In this case sample estimators perform well since we are considering 3958 observations and only 4 assets.

	S&P500	Bond	MM	GSCI
S&P500	1	-0.1216	-0.0667	0.1501
Bond	-0.1216	1	0.2719	-0.0950
MM	-0.0667	0.2719	1	-0.0227
GSCI	0.1501	-0.0950	-0.0227	1

The eigenvalues of the estimated correlation matrix are 0.72, 0.85, 1.06 and 1.38. All eigenvalues are strictly positive and therefore our estimated correlation matrix is positive definite as desired. From these statistics it is clear that bonds and money market present less fluctuations than stocks and commodities, especially for the MM, and are therefore less risky securities. Stocks and commodities have also much lower Sharpe ratios than bonds and MM. We expect then for risk balanced portfolios to be more concentrated in MM and bonds, especially in MM. Another issue to observe in the above statistics is that these assets, except for MM, have a relatively high skewness. This suggests that the use of semi-variance should be more appropriate than variance to describe risk in this case. For every trading day we compute the ERC optimal portfolios $\mathbf{w}_1^{erc}, \ldots, \mathbf{w}_N^{erc}$ solving the following optimization problem for $i = 1, \ldots, N$,

$$\mathbf{w}_i^{erc} = \arg\min f(\mathbf{w}) \quad \text{subject to} \quad \begin{cases} w_j \geq 0, & j = 1, \ldots n \\ \sum_{j=1}^{n} w_j \geq 1 \end{cases} \tag{4.1.1}$$

with

$$f(\mathbf{w}) = K \sum_{j,k=1}^{4} \left(\hat{\rho}_j^i(w) - \hat{\rho}_k^i(w)\right)^2, \tag{4.1.2}$$

where $\hat{\rho}^i_j$ denotes the estimated risk contribution of the instrument j at day i, and $K \in \mathbb{R}$ is a constant. We optimize with long constraint but allowing for arbitrary leverage, the desired leverage can then be obtained normalizing the solution. The factor K in (4.1.2) allows to fasten the optimization: daily risk contributions in this example are pretty small numbers that make the optimization problematic, rescaling with some K big enough helps solving the optimization more efficiently. In this example we used $K = 10^4$ and obtained quite good approximations of the minimum. For the daily portfolio weights of our strategy we can compute using the returns data the daily portfolio returns $\mathbf{x} = (x_1, \ldots, x_N)$ setting $x_i = \mathbf{r}_i' \mathbf{w}_i^{erc}$. Using these N values that depend only on the chosen strategy and on the market data we define the following statistical quantities that we are going to use to measure performance and to compare strategies. We define the mean daily portfolio return by

$$\mu(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^{N} x_i. \tag{4.1.3}$$

The annualized mean portfolio return is then given by

$$\mu(\mathbf{x})_{ann} = \big(1 + \mu(\mathbf{x})\big)^{250} - 1, \tag{4.1.4}$$

where we take the power 250 since there are approximately so many trading days in a year. Note that mean returns are going to be used in order to quantify relationships between portfolio risk and return. To quantify the total gain of the strategy we compute for $k = 1, \ldots, N$ the compounded return,

$$\mu_k^c(\mathbf{x}) = \prod_{i=1}^{k} (1 + x_i) - 1, \tag{4.1.5}$$

and we denote by $\mu^c(\mathbf{x}) = \mu_N^c(\mathbf{x})$ the compounded return over the whole period. As measures of risk of the strategy we compute the sample volatility, value-at-risk and expected shortfall of the daily returns over the period.

$$\begin{aligned} \sigma(\mathbf{x}) &= \left(\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu(\mathbf{x}))^2 \right)^{\frac{1}{2}}, \\ VaR_\alpha(\mathbf{x}) &= -x^{sort}_{\lceil N\alpha \rceil}, \\ ES_\alpha(\mathbf{x}) &= -\frac{1}{\lfloor N\alpha \rfloor} \sum_{i=1}^{\lfloor N\alpha \rfloor} x_i^{sort}. \end{aligned} \tag{4.1.6}$$

These represent daily risks and can be annualized by multiplying them with the factor $\sqrt{250}$. Note that this is the right scaling factor in the case of normally distributed data (see [17] Example 2.23), with no distribution assumption, this is just a definition of how we annualize risks. For the annualized risks we use the notation $\sigma(\mathbf{x})_{ann}$, $VaR_\alpha(\mathbf{x})_{ann}$ and $ES_\alpha(\mathbf{x})_{ann}$. Also very relevant quantities for many asset management firms are the drawdowns of the strategy and in particular the maximum drawdown:

$$\begin{aligned} DD_i(\mathbf{x}) &= \max_{1\le k\le i} \mu_k^c(\mathbf{x}) - \mu_i^c(\mathbf{x}), \\ MDD(\mathbf{x}) &= \max_{1\le i\le N} DD_i(\mathbf{x}). \end{aligned} \tag{4.1.7}$$

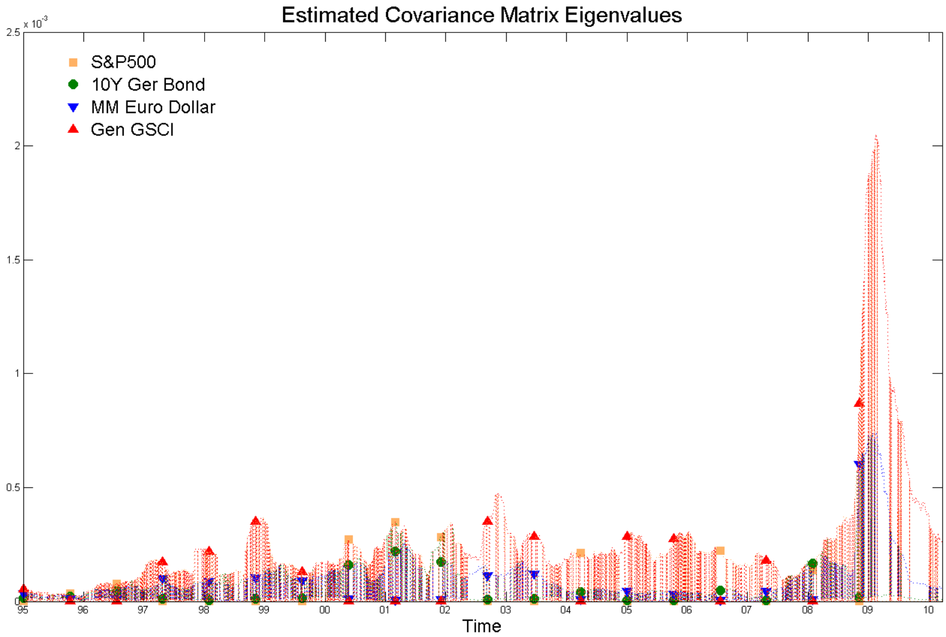
We will then consider the following performance ratios: $S_\sigma(\mathbf{x}) = \frac{\mu(\mathbf{x})_{ann}}{\sigma(\mathbf{x})_{ann}}$, $S_{VaR}(\mathbf{x}) = \frac{\mu(\mathbf{x})_{ann}}{VaR_\alpha(\mathbf{x})_{ann}}$, $S_{ES}(\mathbf{x}) = \frac{\mu(\mathbf{x})_{ann}}{ES_\alpha(\mathbf{x})_{ann}}$, $S_{DD}(\mathbf{x}) = \frac{\mu(\mathbf{x})}{\frac{1}{N}\sum_{i=1}^N DD_i(\mathbf{x})}$ and $S_{MDD}(\mathbf{x}) = \frac{\mu(\mathbf{x})}{MDD(\mathbf{x})}$. Another very important quantity in practice is the portion we need to trade in order to apply a certain investment strategy. To quantify this we define the strategy *turnover* as

$$TO_i(\mathbf{w}_1, \ldots, \mathbf{w}_N) = \sum_{k=1}^{i-1} |\mathbf{w}_{k+1} - \mathbf{w}_k|_1 \tag{4.1.8}$$

where $|\mathbf{w}|_1 = |w_1| + \cdots + |w_n|$ denotes the 1-norm, and we denote the terminal turnover by $TO(\mathbf{w}_1, \ldots, \mathbf{w}_N) = TO_N(\mathbf{w}_1, \ldots, \mathbf{w}_N)$. The mean daily turnover can be annualized by multiplying it with 250. We now expose our results for the ERC strategy using different risk measures and different distribution assumptions. We have implemented everything using Matlab 7.8.0. Our implementation consists essentially in using the Matlab function *fmincon* (see Appendix C) properly to solve the optimization.

4.1.1 Standard Deviation

We consider the ERC strategy using the standard deviation as risk measure. For every trading day we estimate the covariance matrix of the returns distribution using the previous 100 daily returns and the shrinkage procedure of Section 3.2 selecting as prior a covariance matrix with constant correlation coefficient. For this example the difference between shrinkage estimation and sample covariances is however not so pronounced since we are considering 100 data per instrument and only 4 instruments. The eigenvalues of the estimated covariance matrix for every trading day are positive and represented in the following figure.



The daily estimated risk contributions in (4.1.2) can then be computed using the daily estimated covariance matrices and relation (2.2.1). We also compute the $1/n$ and MV portfolios and compare the results. We do not consider leveraged positions for now.

	MV	ERC	$1/n$
μ	0.003%	0.005%	0.014%
μ_{ann}	0.70%	1.19%	3.63%
μ^c	11.62%	20.54%	66.77%
σ	0.040%	0.071%	0.516%
$VaR_{0.01}$	0.10%	0.19%	1.34%
$ES_{0.01}$	0.16%	0.27%	2.11%
σ_{ann}	0.64%	1.12%	8.16%
VaR_{ann}	1.63%	3.02%	21.17%
ES_{ann}	2.55%	4.27%	33.37%
mean DD	0.51%	0.52%	9.51%
MDD	2.17%	4.01%	64.38%
S_σ	1.09	1.06	0.44
S_{VaR}	0.43	0.39	0.17
S_{ES}	0.27	0.28	0.11
S_{DD}	0.55%	0.91%	0.15%
S_{MDD}	0.13%	0.12%	0.02%
ann. mean TO	0.41	0.97	0

Figure 4.1: Paths for the 3 strategies in the 4 assets example.

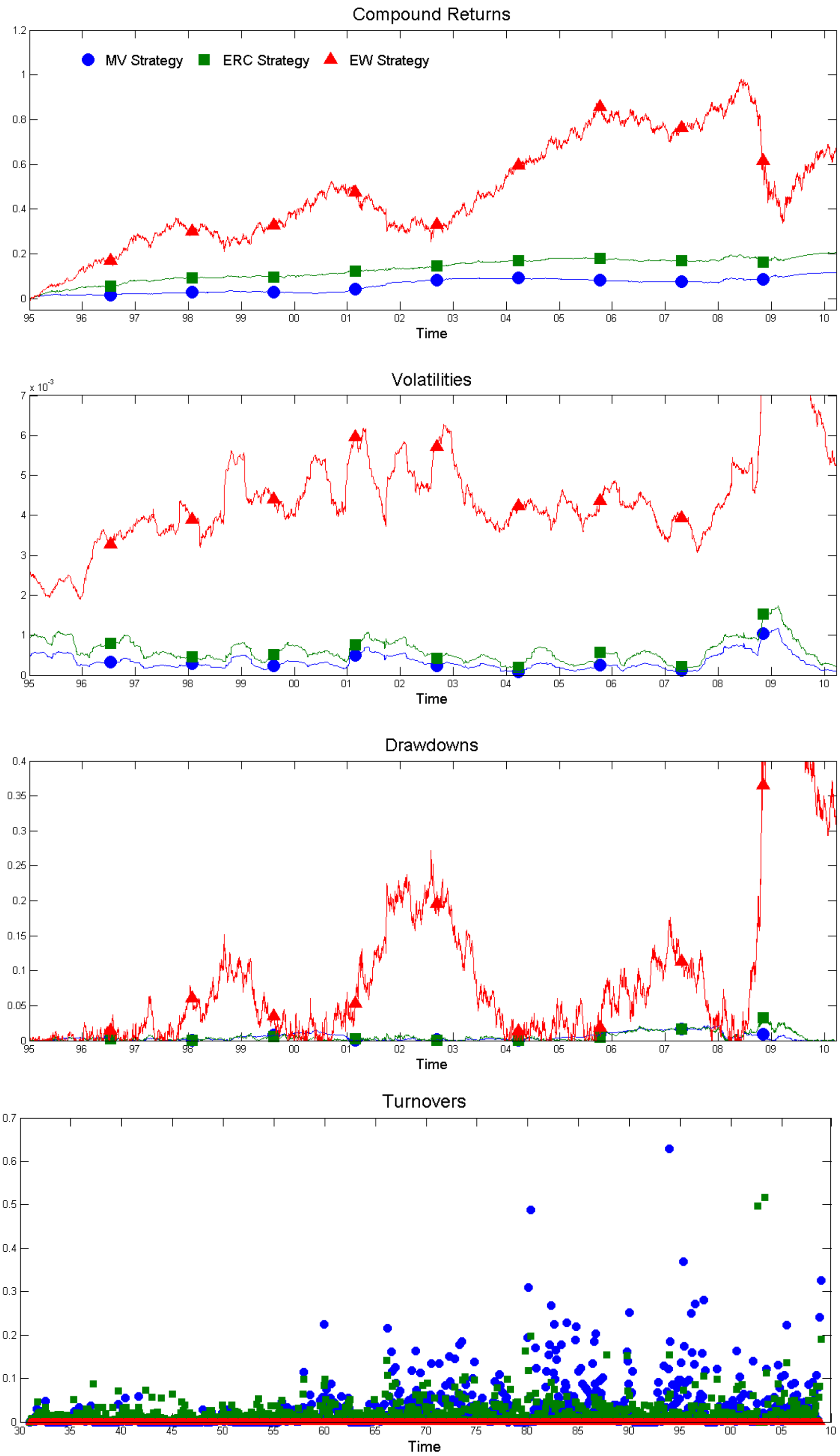
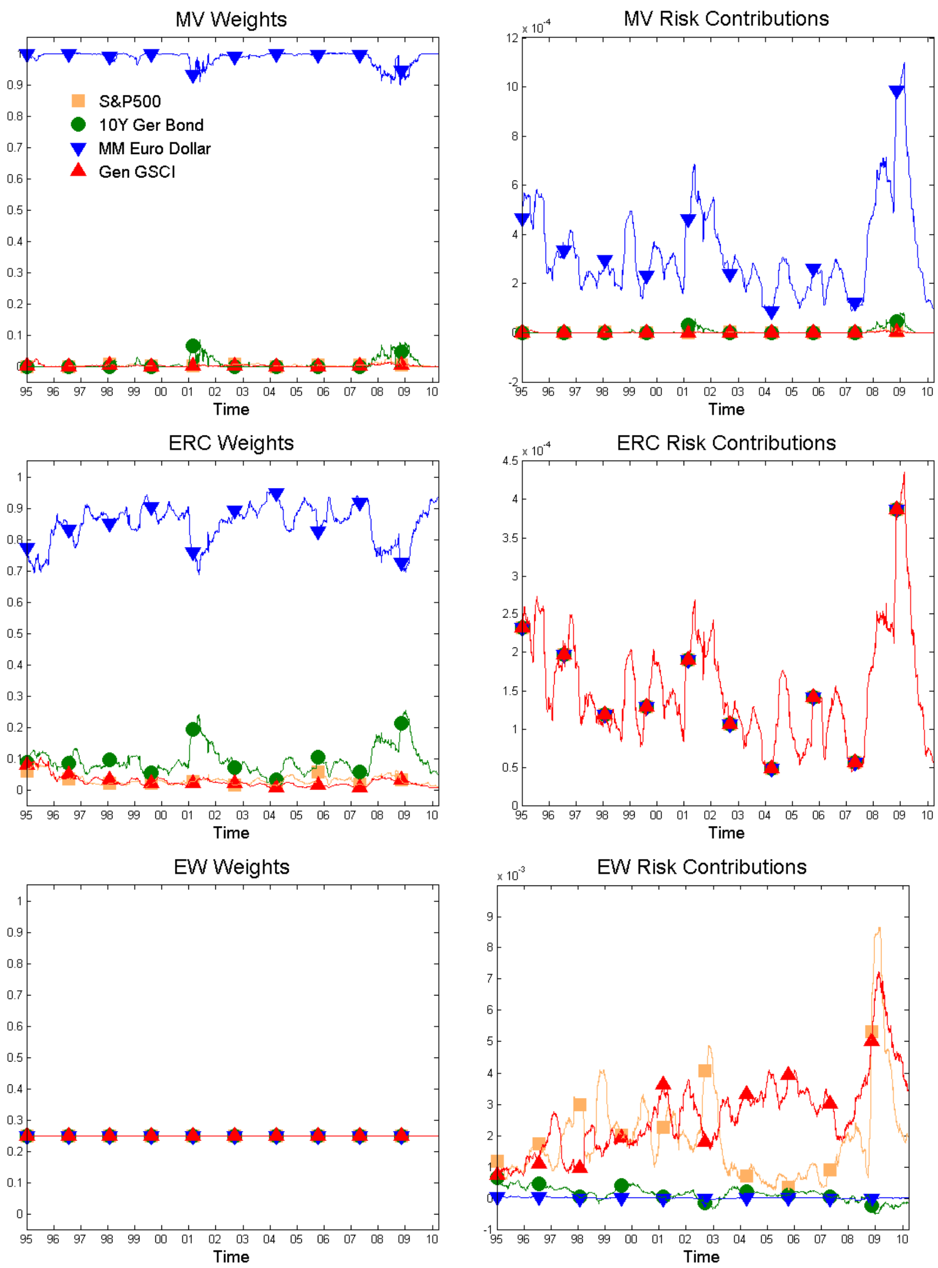
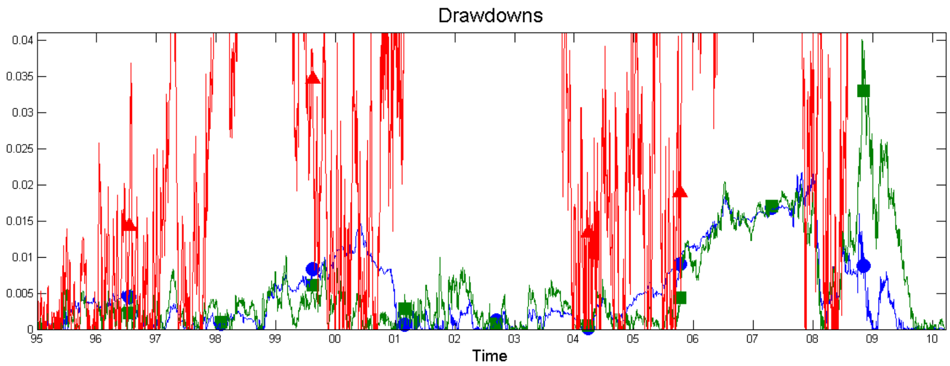


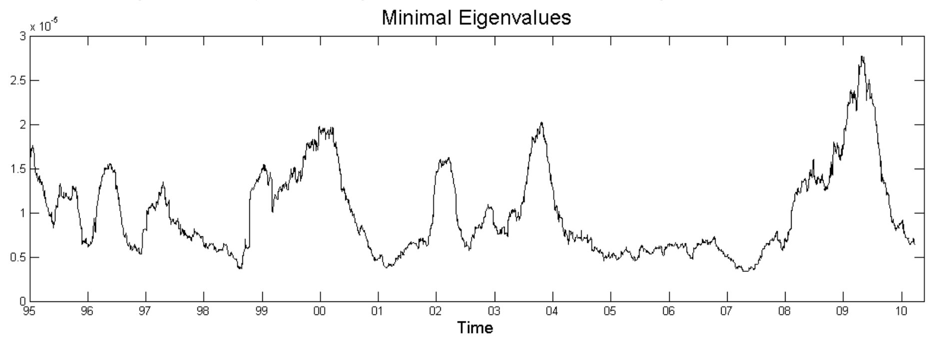
Figure 4.2: Weights and risk contributions of the 4 assets.



As we could already say at the beginning by looking at the statistics of the instruments MV and ERC portfolios are concentrated very strongly in Money Market, some bonds and almost no exposition in stocks and commodities. This is particulary accentuated in the MV strategy where the position in stocks and commodities is often essentially zero and always under 3%, see Figure 4.2. This is because of the huge difference in risks among these assets. If we compare the results looking in the above table we can observe that MV and ERC strategies exhibit much better performance ratios than splitting the wealth equally. In Figure 4.1 we can observe that the ERC strategy is located between MV and EW in terms of risk measures, returns and drawdowns with similar statistics as the MV portfolio. We bring the drawdowns again using another scale to compare MV and ERC portfolios.



Looking at the portfolio weights and turnovers plots we can observe that MV and ERC positions are quite stable in time, especially for MV, where almost all the exposures remain in the MM. The MV strategy exhibits better turnover than ERC. The situation becomes much more dynamic if we eliminate the MM from the portfolio and consider only stocks, bonds and commodities. The eigenvalues of the estimated covariance matrices in the case of 3 assets are always positive as we can see from the following graph representing for every trading period the minimal eigenvalue.



Doing the computations again without MM we observe the following.

	MV	ERC	$1/n$
μ	0.017%	0.019%	0.018
μ_{ann}	4.42%	4.99%	4.63%
μ^c	95.15%	1.11	86.33%
σ	0.29%	0.34%	0.69%
$VaR_{0.01}$	0.80%	0.86%	1.77%
$ES_{0.01}$	0.96%	1.11%	2.82%
σ_{ann}	4.58%	5.44%	10.88%
VaR_{ann}	12.62%	13.58%	27.96%
ES_{ann}	15.19%	17.54%	44.62%
mean DD	3.95%	4.78%	15.21%
MDD	20.58%	26.92%	99.06%
S_σ	0.97	0.92	0.43
S_{VaR}	0.35	0.37	0.17
S_{ES}	0.29	0.28	0.10
S_{DD}	0.44%	0.41%	0.12%
S_{MDD}	0.084%	0.072%	0.018%
ann. mean TO	2.13	1.52	0

As in the previous case we again observe that MV and ERC strategies perform much better than EW in terms of Sharpe ratio and drawdowns. Similar plots as in the previous example are given in Figure 4.3 and 4.4. One important difference with respect to the previous situation with 4 assets is that ERC is not only better than EW in terms of performance ratio, but it also has a higher terminal compound return. Looking at the turnovers we can also observe that eliminating the MM we need to trade much more to keep the risks balanced. In this example ERC has better turnover than MV. Going back to our initial example with 4 assets, a possible way to avoid high concentrated portfolios in the MM, is to add weights constraints to the optimization problem. We tried to optimize the risk contributions as in the ERC approach setting the additional constraints $w_3 \leq 0.5, 0.6, 0.7$, which guarantees that no more than 50%, 60% or 70% respectively of our wealth is invested in the MM, and obtained the paths of Figure 4.5.

Figure 4.3: Paths for the 3 strategies without Money Market.

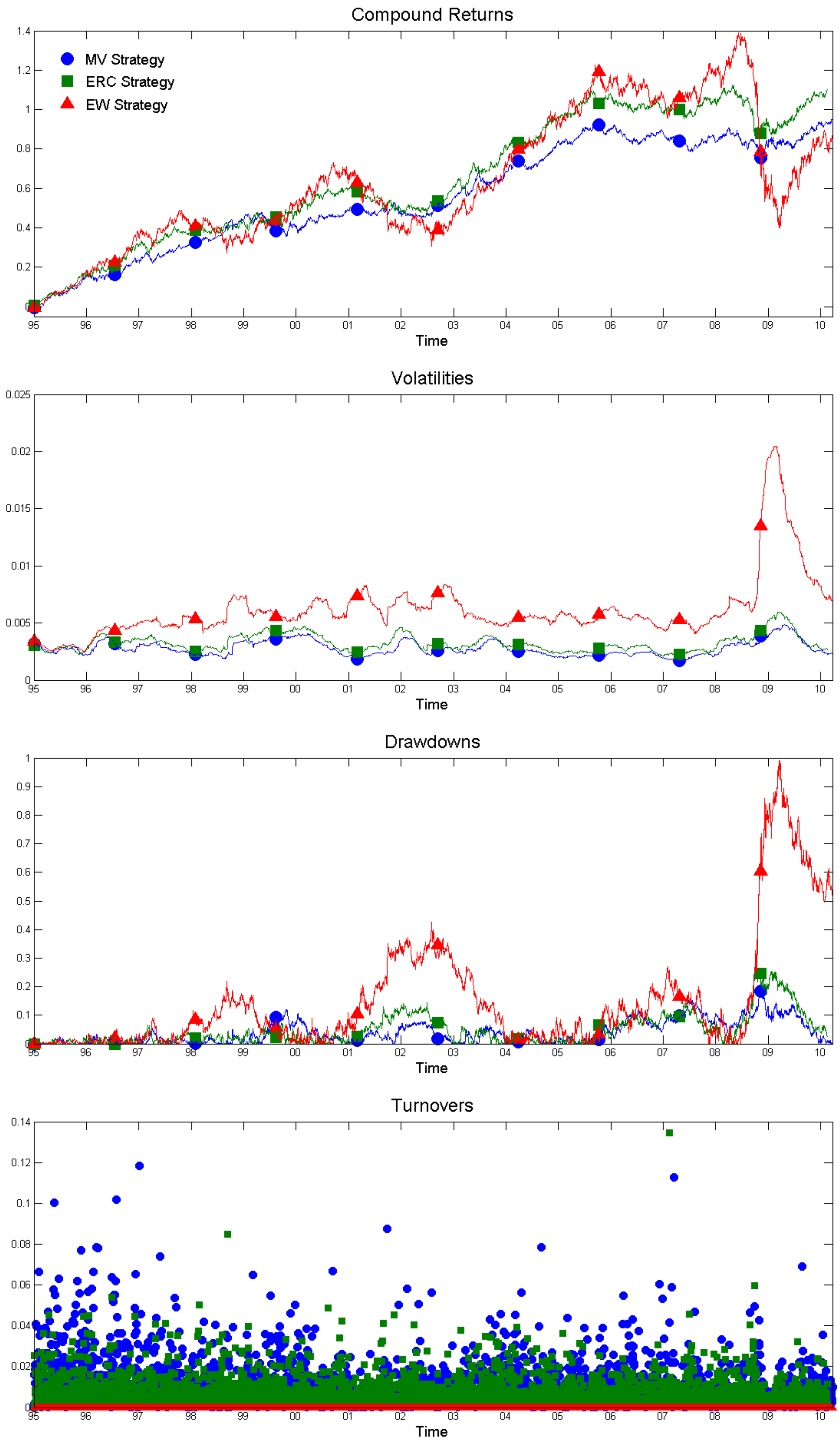


Figure 4.4: Weights and risk contributions of the 3 assets.

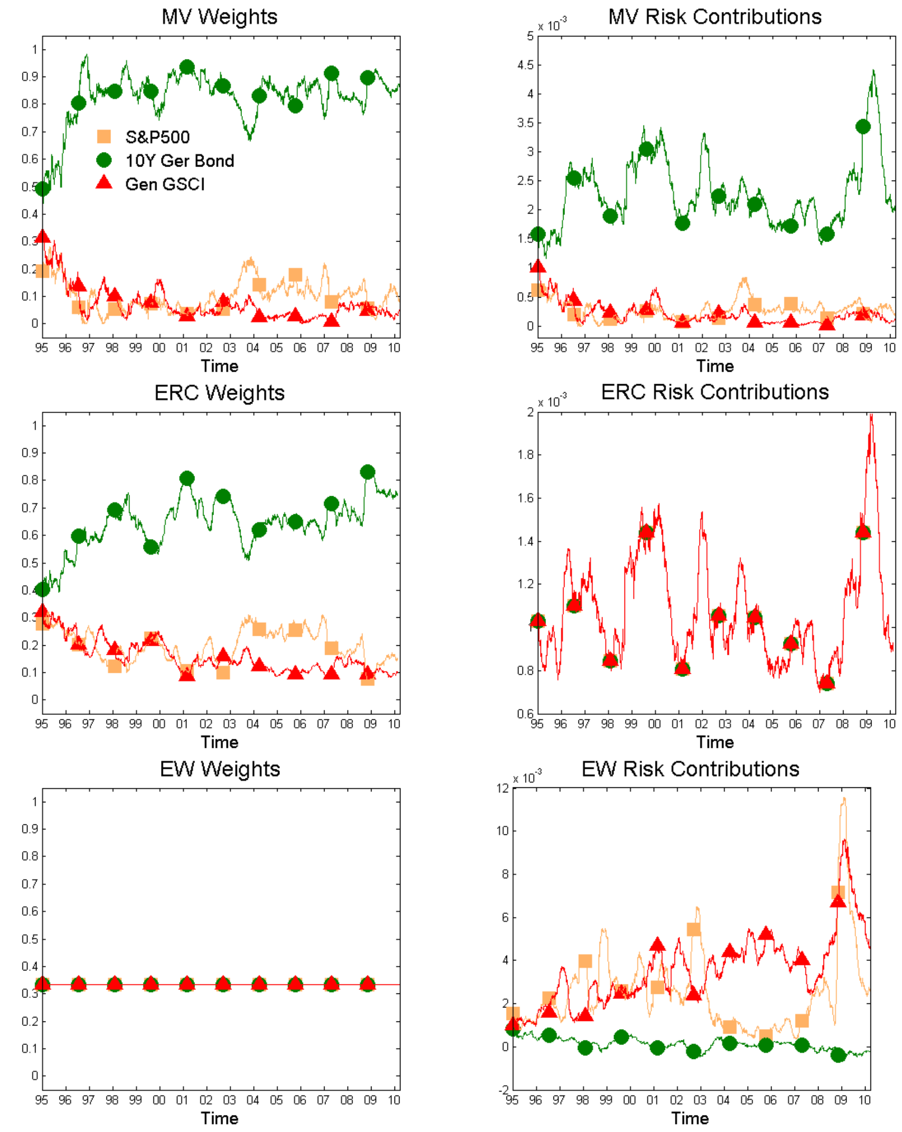
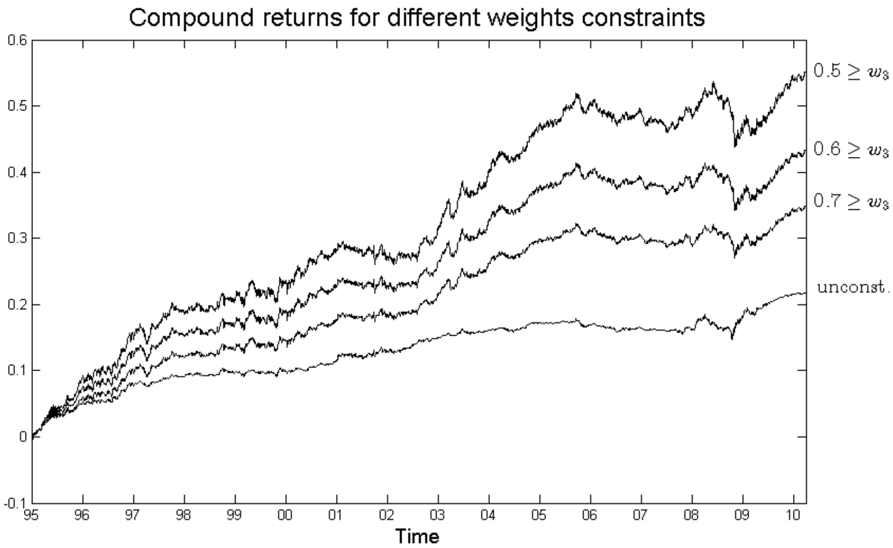


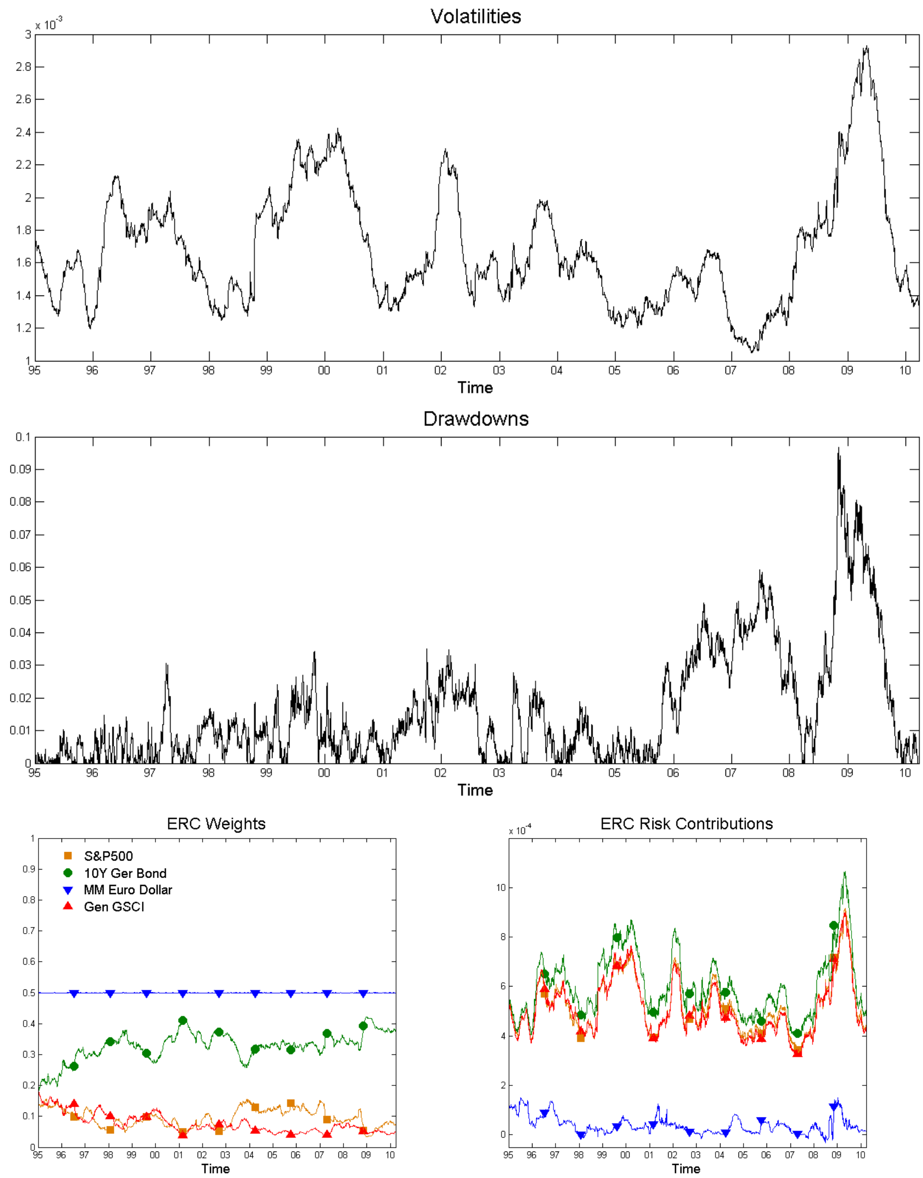
Figure 4.5: Using weights constraints.



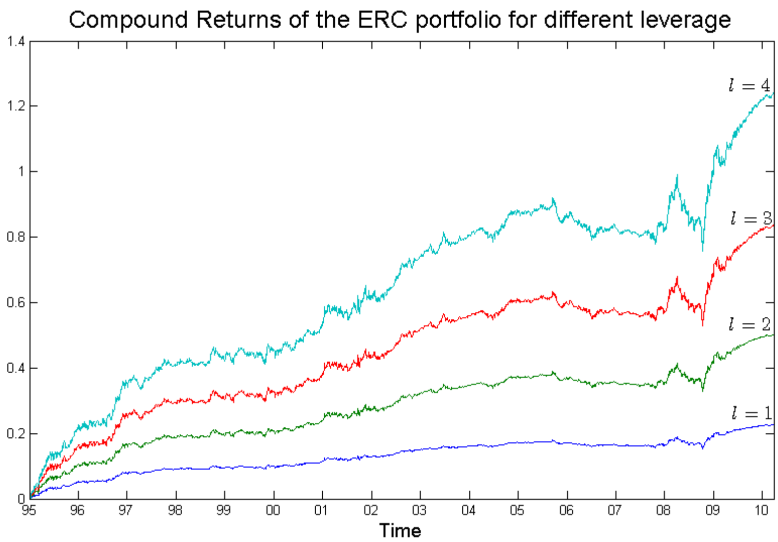
	unconstraint	0.7	0.6	0.5
μ	0.005%	0.008%	0.009%	0.011%
μ_{ann}	1.19%	1.94%	2.36%	2.78%
μ^c	20.54%	35.18%	44.13%	53.52%
σ	0.071%	0.11%	0.14%	0.17%
$VaR_{0.01}$	0.19%	0.28%	0.36%	0.44%
$ES_{0.01}$	0.27%	0.35%	0.45%	0.55%
σ_{ann}	1.12%	1.73%	2.23%	2.73%
VaR_{ann}	3.02%	4.36%	5.67%	6.96%
ES_{ann}	4.27%	5.47%	7.05%	8.68%
mean DD	0.52%	0.85%	1.17%	1.57%
MDD	4.01%	5.18%	7.25%	9.67%
S_σ	1.06	1.12	1.06	1.02
S_{VaR}	0.39	0.44	0.42	0.4
S_{ES}	0.28	0.35	0.34	0.32
S_{DD}	0.91%	0.90%	0.80%	0.70%
S_{MDD}	0.12%	0.15%	0.13%	0.11%
ann. mean TO	97%	49.34%	63.77%	79.19%

Here we can observe something very interesting: imposing constraint to the money invested in the MM, we get portfolios with higher returns and higher risks and drawdowns, however without affecting the performance ratios strongly. If we consider for example a 50% constraint in the MM we obtain a twice as big mean daily return with respect to the unconstraint ERC solution. Additionally, we observe that the Sharpe ratio remains very

Figure 4.6: Some diagrams in the case of 50% constraint in Money Market.



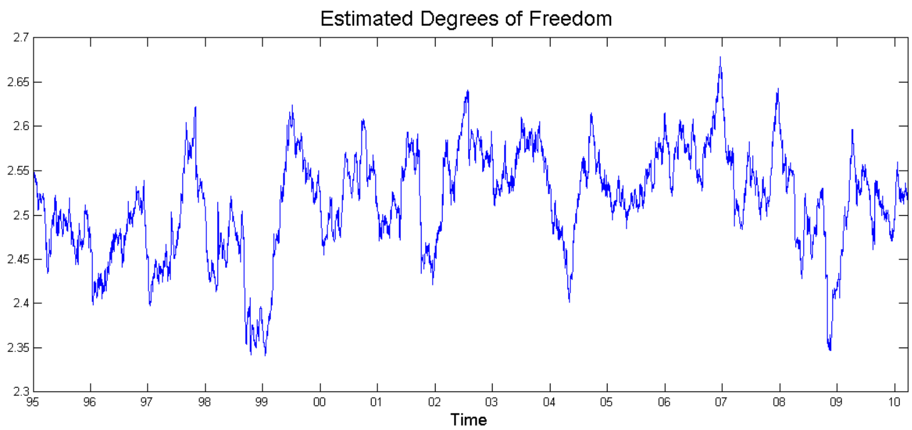
close to the one of the unconstraint solution. In other words what we observe here is that computing equal contributions to risk with constraints on the MM weight allows to reach some return target we could not reach with the unconstraint ERC. The strategy has lower turnovers with weights constraints since the position in the MM is typically constant and equal to the upper bound. The performance measures we have introduced so far may appear meaningless looking at these numerical results: we are optimizing the allocation under stronger conditions in the portfolio weights and we may find portfolios with higher performance measures. This is because in the ERC approach we are not optimizing the portfolio with respect to these performance measures rather equalizing risk contributions. We are going to discuss the problem of measuring diversification in Chapter 5. Another way to obtain returns increase is to consider leveraged portfolios. The use of leverage is quite common in practice especially when trading futures. We will explain more precisely how leverage can be used in practice to reach certain return or risk targets in Section 4.2. We briefly illustrate the leverage effect here considering fixed daily leverages of $l = 1, 2, 3, 4$. Doing this, we obtained the following compound returns paths.



Using leverage we observe a similar effect as in the case of MM weight constraints: return are increased but also fluctuations. Another desirible feature of the strategy for practitioners is that the obtained portfolio weights should not be too sensitive in the estimated covariances. Since estimating covariance matrix is not a trivial issue as discussed in Chapter 3, if small changes in the covariance matrix entries lead to very different portfolio weights, then the method cannot be applied in practice. We analyzed this empirically for this example comparing the results using different number of observations to estimate the covariance matrices at each trading day. We observed analogous results as in the next example (see Section 4.2.4).

4.1.2 Normal Assumption vs. t Assumption

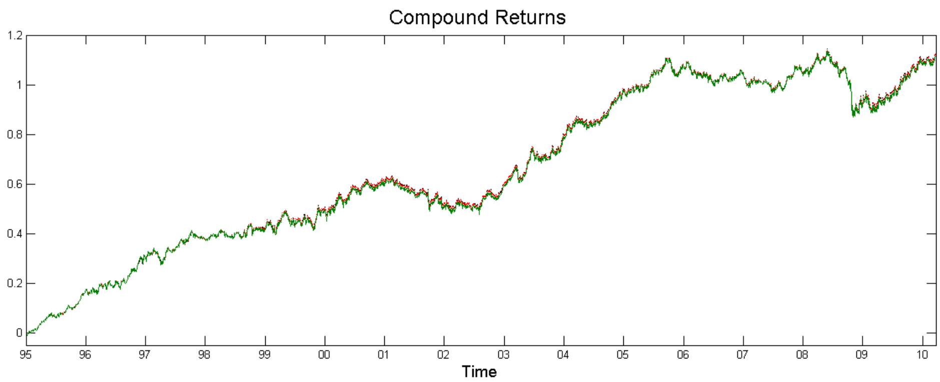
In this section we want to consider ERC principles for VaR and ES assuming normally or t-distributed data as in the Examples 1.1.1 and 1.1.4. We exclude the Money Market from the portfolio and consider only stocks, bonds and commodities. For every trading day we estimate the covariance matrix and the expected returns using 100 daily observations. For the covariance matrix we use as before the shrinkage estimator and for the expected returns we use the sample mean. We then estimate risk contributions using formula (2.3.2). Not unproblematic is how to fit the parameter ν for the t-distribution. For every trading period we use the *EM algorithm* (expectation maximization), see [17] Section 3.2.4. Testing this algorithm on this example we have observed that it converges rapidly. We used for every trading period 100 observations and 50 iterations of this procedure to estimate ν. The results are represented in the following diagram.



Note that the closer ν is to the value 2, the larger are the tails. We summarize the portfolios statistics for a confidence level $\alpha = 0.01$ in the following table.

	VaR normal	VaR t	ES normal	ES t
μ	0.020%	0.021%	0.021%	0.021%
μ_{ann}	5.18%	5.38%	5.31%	5.38%
μ^c	1.17	1.24	1.22	1.24
σ	0.34%	0.34%	0.35%	0.34%
$VaR_{0.01}$	0.86%	0.86%	0.86%	0.86%
$ES_{0.01}$	1.11%	1.10%	1.12%	1.12%
σ_{ann}	5.44%	5.44%	5.46%	5.45%
VaR_{ann}	13.57%	13.54%	13.61%	13.53%
ES_{ann}	17.60%	17.43%	17.74%	17.70%
mean DD	4.44%	3.99%	4.20%	4.14%
MDD	26.92%	26.10%	27.21%	26.95%
S_σ	0.95	0.99	0.97	0.99
S_{VaR}	0.38	0.38	0.38	0.38
S_{ES}	0.29	0.30	0.29	0.29
S_{DD}	0.46%	0.53%	0.49%	0.51%
S_{MDD}	0.075%	0.080%	0.076%	0.077%

The strategy statistics are almost identical, and the compound return paths are also very similar. The weights and risk contributions for the two distribution assumptions are represented in Figure 4.7 and 4.8.



We also represent portfolio weights and risk contributions in each case.
In this example we can observe that using VaR or ES under normal or t distribution assumption do not cause essentially any difference in the portfolio statistics or in the portfolio weights. If we compare these results with the ERC portfolio of Section 4.1.1 using the volatility as risk measure we do not observe any pronounced difference in performances or drawdowns.

4.1.3 VaR vs. ES

We now compute ERC strategies estimating daily risk contributions of VaR and ES using only the historical data without any distribution assumption.

Figure 4.7: Weights and risk contributions using the normal distribution.

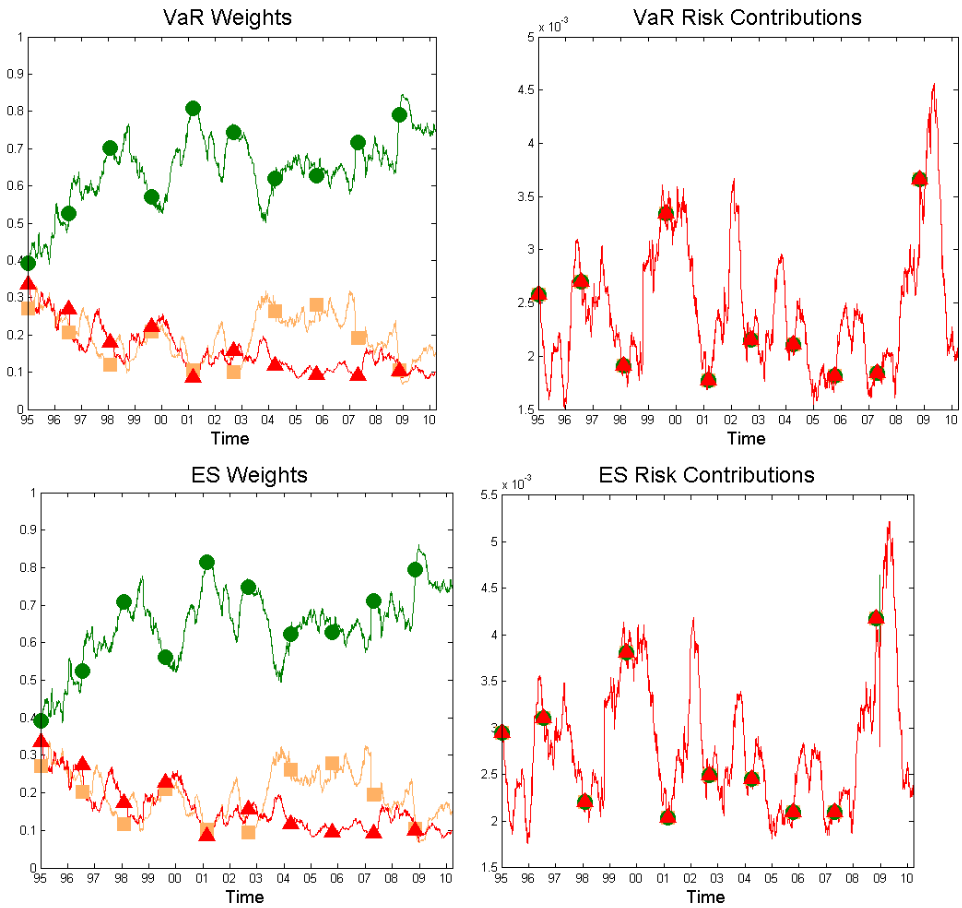
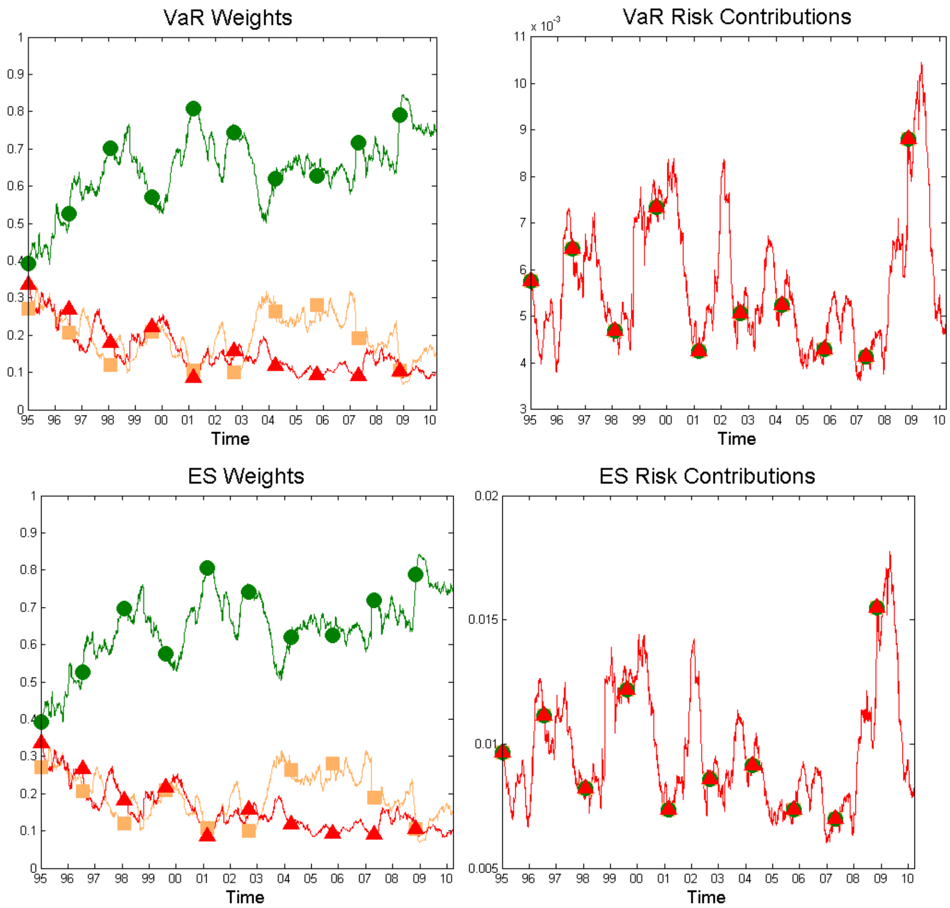


Figure 4.8: Weights and risk contributions using the t-distribution.

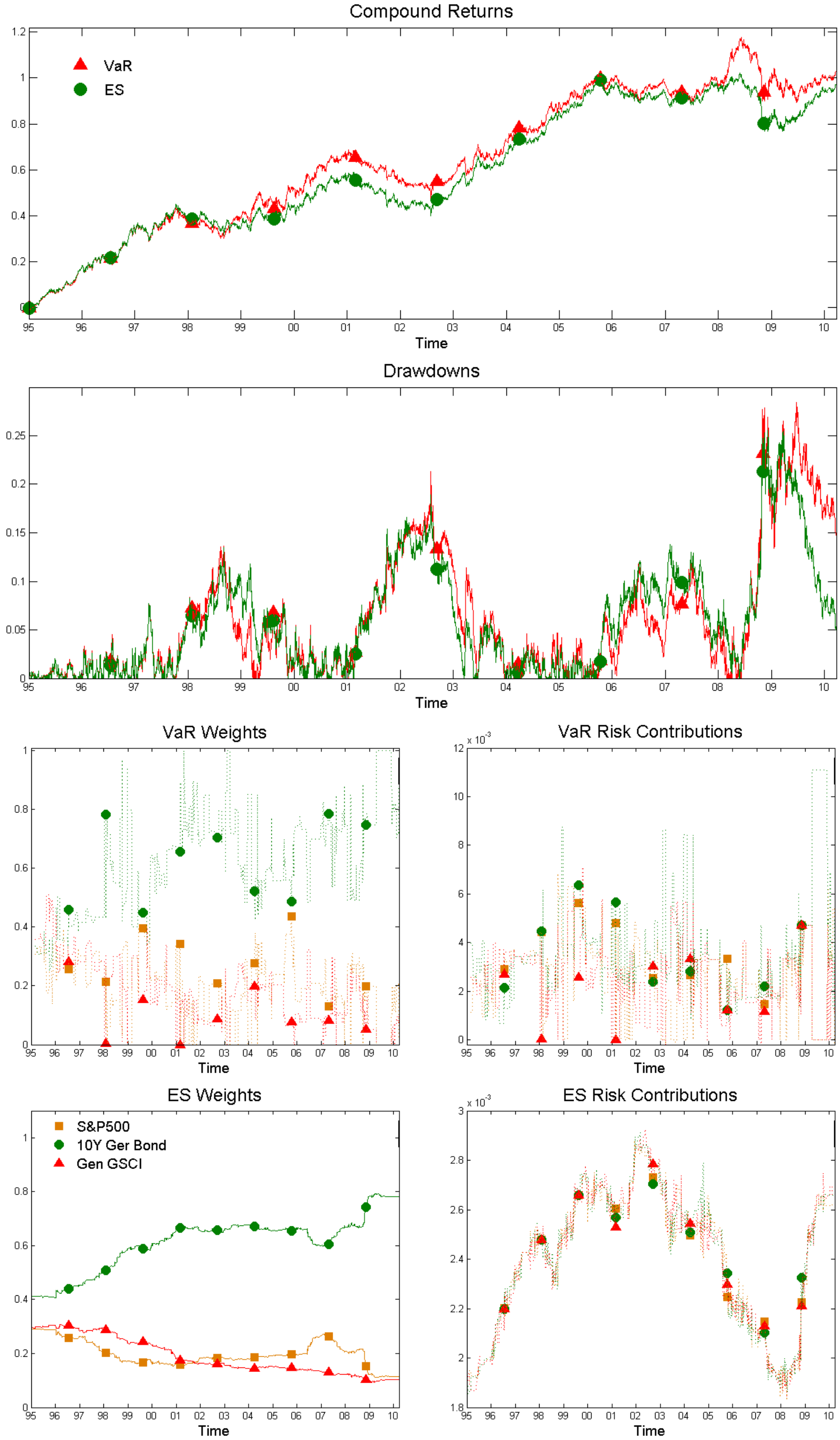


As in the previous section we exclude the MM form the portfolio. For every trading day we consider the previous 1000 daily returns observations to estimate the risk contributions. For the case of the expected shortfall the estimation is quite straight forward as in (3.2.4) with quite accurate results. For the case of VaR we have used the kernel estimation (3.2.5) with bandwidth $b = 10^{-12}$, which was verified to give satisfactory approximations.

	VaR	ES
μ	0.018%	0.018%
μ_{ann}	4.78%	4.57%
μ^c	1.03	97.24%
σ	0.4%	0.37%
$VaR_{0.01}$	1.00%	0.91%
$ES_{0.01}$	1.29%	1.17%
σ_{ann}	6.29%	5.91%
VaR_{ann}	15.87%	14.47%
ES_{ann}	20.32%	18.45%
mean DD	6.12%	5.88%
MDD	28.42%	25.60%
S_σ	0.76	0.77
S_{VaR}	0.30	0.32
S_{ES}	0.24	0.25
S_{DD}	0.0031%	0.30%
S_{MDD}	0.066%	0.070%
ann. mean TO	3.13	0.23

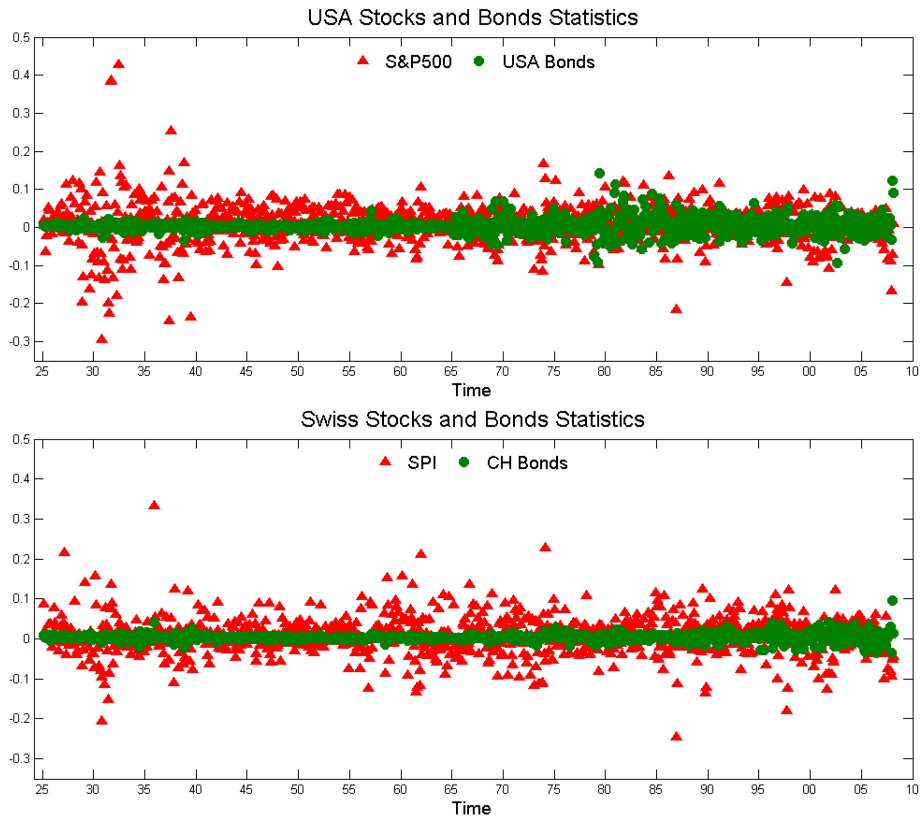
In this case it is quite hard to equalize the risk contributions since the estimator (3.2.4) and (3.2.5) are not smooth functions of the portfolio weights. In the case of ES the estimation is much easier to do and much faster than for VaR. ES shortfall lead to more stable strategies than VaR. The annualized mean turnover of the VaR strategy is more than 10 times bigger than the one of the ES strategy. The performance ratios in this case are lower than the ones using standard deviation of Section 4.1.1 or the ones using normal or t assumption of Section 4.1.2. Some statistics are represented in Figure 4.9.

Figure 4.9: Some diagrams for ERC strategies constructed with VaR and ES risk contributions estimated using the empirical distributions.



4.2 Stocks & Bonds Portfolios

In this example we consider portfolios of USA/CH stocks and bonds over a very long period of more or less 77 years. We consider the period from January 1931 to december 2008 using monthly observations and rebalancing the portfolio monthly. The stocks and bonds statistics for these two countries from the 20s are the following.



	S&P500	USA Bonds	SPI	CH Bonds
Mean	0.92%	0.47%	0.70%	0.37%
Ann. Mean	11.62%	5.85%	8.77%	4.54%
Max.	42.86%	14.35%	33.35%	9.62%
Min.	-29.45%	-9.22%	-24.62%	-3.57%
Standard deviation	5.53%	1.99%	4.74%	0.91%
Ann. standard deviation	19.17%	6.88%	16.43%	3.15%
Sharpe	0.61	0.85	0.53	1.44
Skewness	37.89%	97.12%	20.45%	92.46%
Kurtosis	12.75	10.28	7.84	16.67

The sample correlations are given by the following table.

	S&P500	USA Bonds
S&P500	1	0.0993
USA Bonds	0.0993	1
	SPI	CH Bonds
SPI	1	0.1323
CH Bonds	0.1323	1

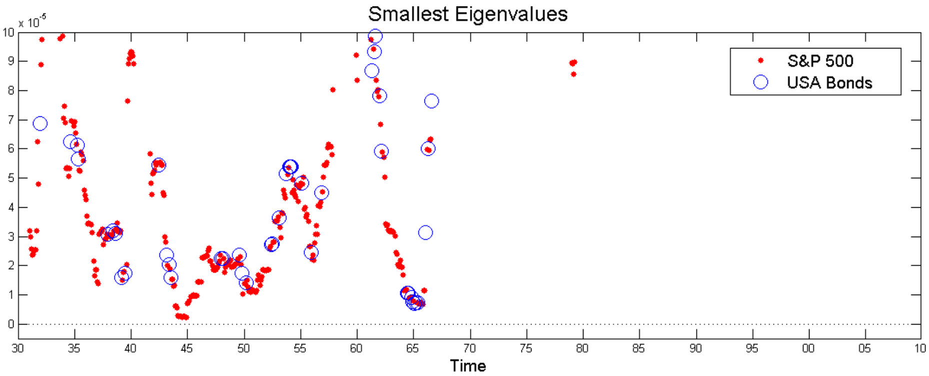
Considering all the 4 assets together we find the following sample correlations.

	S&P500	USA Bonds	SPI	CH Bonds
S&P500	1	0.0993	0.4632	-0.0034
USA Bonds	0.0993	1	0.0245	0.3521
SPI	0.4632	0.0245	1	0.1323
CH Bonds	-0.0034	0.3521	0.1323	1

The eigenvalues of the last correlation matrix are 0.47, 0.71, 1.27 and 1.55. These three correlation matrices are positive definite as desired. From these statistics we can observe that american stocks have performed better than swiss stocks and that swiss bonds have performed better than the american ones in terms of Sharpe ratios. Stocks present as in the previous example much more fluctuations than bonds and therefore we expect for risk balanced portfolios to be more concentrated in bonds. To observe is also that bonds have high positive skewness. Swiss and american stocks are strongly correlated between each other and the same holds for bonds, whereas correlations between stocks and bonds are lower. Using the standard deviation we construct Risk Parity Portfolios and analyze the results. For the covariance matrix estimation we use the shrinkage estimator of Section 3.1 every month using the previous 24 monthly observations. In this example, for graphical representation purposes, we define cumulative returns by summing up the monthly portfolio returns instead of compounding them geometrically as in Section 4.1. This is because, over such a long time period, by geometrically compounded returns, the fluctuations in the first years are not visible in the graphical representation. We define therefore,

$$\mu_k^c(\mathbf{x}) = \sum_{i=1}^{k} x_i. \tag{4.2.1}$$

This corresponds to the situation in which after every trading period the portfolio balance is established again to the initial wealth before the next trading. Whereas by geometrically compounded returns represent the evolution of the capital without introducing or withdrawing money. Drawdowns are computed using the paths of the cumulative returns computed as in (4.2.1). In the introduction of this thesis we considered the example of stocks and bonds portfolios to explain the principle of risk parity. We now want to construct equal contributions to risk portfolios and see if risk diversification in terms of risk contributions really provides benefits in terms of performance as hinted by Qian in [22]. Before computing the ERC portfolios we check the eigenvalues of the estimated covariance matrices to be strictly positive. This is the case. We represent the smallest eigenvalues for the USA asset returns covariance matrices in the following diagram.



4.2.1 USA Risk Parity Strategies

We expose the portfolio statistics, returns, drawdowns, volatilities and turnover diagrams of the ERC strategy and the usual benchmarks. We also bring diagrams representing portfolio weights and risk contributions ratios for the three different strategies as we did in the previous example.

	MV	ERC	$1/n$
μ	0.50%	0.56%	0.70%
μ_{ann}	6.21%	6.93%	8.71%
μ^c	4.70	5.24	6.53
σ	1.88%	1.96%	3.04%
$VaR_{0.01}$	4.29%	4.56%	7.01%
$ES_{0.01}$	5.98%	5.99%	10.49%
σ_{ann}	6.50%	6.79%	10.52%
VaR_{ann}	14.87%	15.79%	24.28%
ES_{ann}	20.73%	20.74%	36.35%
mean DD	1.26%	1.39%	3.21%
MDD	19.13%	15.17%	56.31%
S_σ	0.95	1.02	0.83
S_{VaR}	0.42	0.44	0.36
S_{ES}	0.30	0.33	0.24
S_{DD}	0.40	0.40	0.22
S_{MDD}	0.026	0.037	0.012
ann. mean TO	35.68%	23.35%	0

Figure 4.10: Paths for MV, ERC and $1/n$ strategies of USA stocks and bonds.

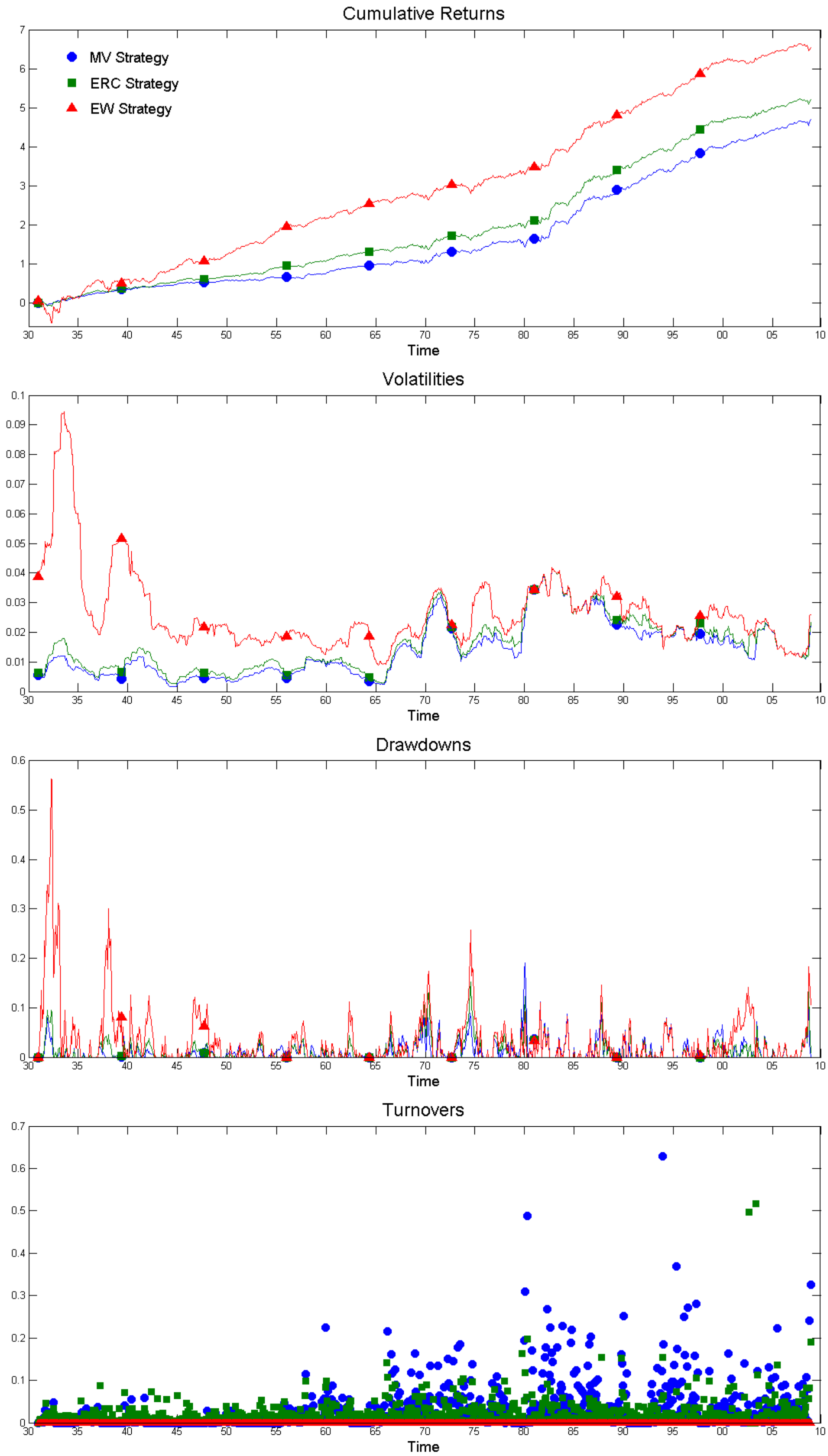
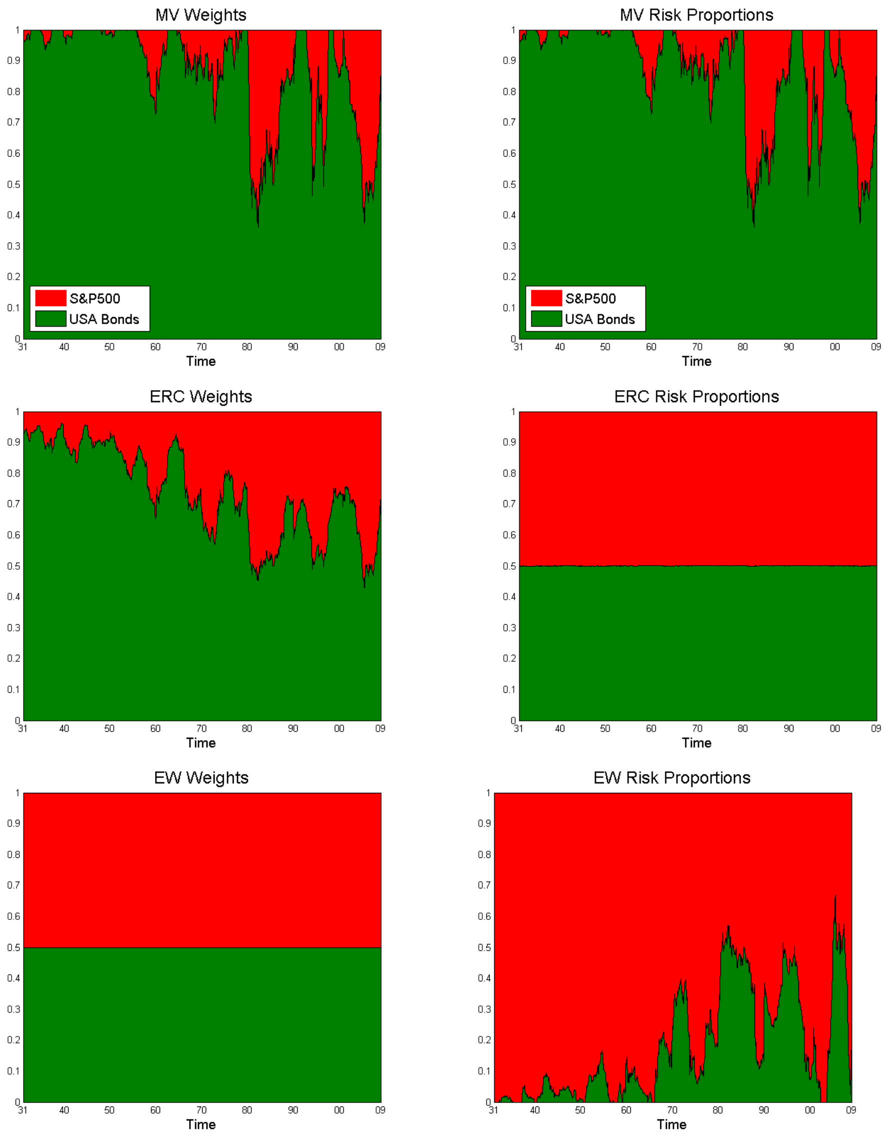


Figure 4.11: USA stocks/bonds weights and risk proportions.



The performance benefits in terms of Sharpe ratios that Qian [22] claimed to have observed are exactly visible also in this example. The Sharpe ratios of USA stocks and bonds are respectively 0.61 and 0.85. The 50/50 strategy has a Sharpe ratio of 0.83, which is smaller than the one of bonds and this means that there is no benefit in splitting the wealth among stocks and bonds. This means that the EW strategy has poor diversification. The ERC strategy provides higher Sharpe ratio than stocks and bonds individually and this represent the benefits of diversification. The Sharpe ratio of the MV portfolio is also higher than the ones of the assets. In particular we can observe that ERC present the highest Sharpe ratio of 1.02 among the three strategies. Performance ratios for the VaR and ES risk measures are also better for the ERC strategy. It is important to observe that the 50/50 strategy suffers from dramatic drawdowns compared with MV and ERC, MV has the best mean drawdown and ERC has the best maximum drawdown over the whole period. A stable strategy in terms of drawdowns is very important for asset management firms. Looking at the risk proportions diagrams we can observe that the risk contributions optimization has been performed successfully. The risk contributions are very precisely equalized. Interesting is to observe that by MV the risk proportion and the portfolio weights seem to be exactly equal. This is indeed true and can be easily verified analytically. The MV portfolio minimizes the variance $\mathbf{w}'\Omega\mathbf{w}$ under the constraint $\mathbf{w}'\mathbf{1} = 1$, and hence must satisfy for $i = 1, \ldots, n$ and some Lagrange multiplier λ

$$(\Omega\mathbf{w})_i - \lambda = 0. \tag{4.2.2}$$

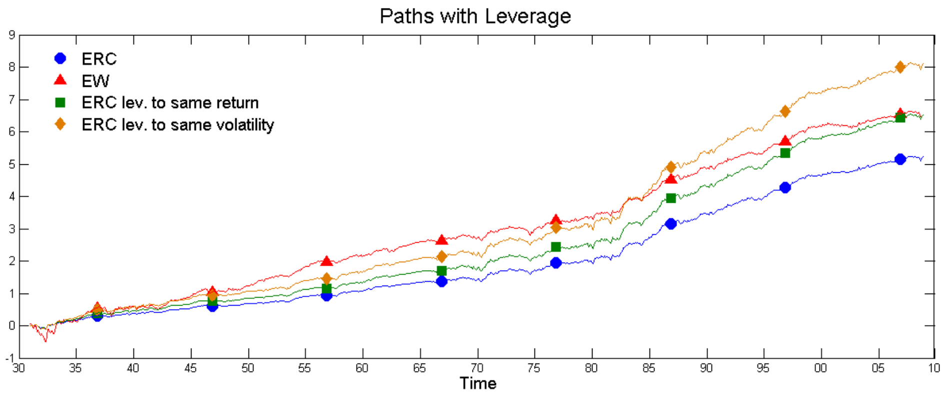
The risk proportions can be written as

$$\frac{RC_i(\mathbf{w})}{\sqrt{\mathbf{w}'\Omega\mathbf{w}}} = \frac{w_i(\Omega\mathbf{w})_i}{\mathbf{w}'\Omega\mathbf{w}} = \frac{\lambda}{\mathbf{w}'\Omega\mathbf{w}} w_i. \tag{4.2.3}$$

Using the constraint it follows

$$1 = \sum_{i=1}^{n} \frac{RC_i(\mathbf{w})}{\sqrt{\mathbf{w}'\Omega\mathbf{w}}} = \frac{\lambda}{\mathbf{w}'\Omega\mathbf{w}} \sum_{i=1}^{n} w_i = \frac{\lambda}{\mathbf{w}'\Omega\mathbf{w}}, \tag{4.2.4}$$

and hence combining the last two equations we see that for the strategy that minimizes the variance the risk ratios must be the same as the portfolio weights. This suggests an alternative interpretation of how the strategy are constructed: by MV the risk is partitioned in the same way as the capital, by ERC the risk is partitioned equally and by EW the capital is partitioned equally. In terms of cumulative returns ERC is located between MV and EW, and the same holds for the volatility confirming the result of Theorem 2.2.3. The ERC weights are stable in time, this is also very desirable in practice, since high turnovers involve high transaction costs. The ERC portfolio exhibits a quite low turnover with an annualized mean of 23.35%. While ERC has high Sharpe ratio, it has lower return than EW because of the much lower volatility. An investor may therefore not be able to achieve a certain return objective just by constructing a strategy with high Sharpe ratio. A possible way of dealing with low return is to leverage the positions in order to achieve a higher level of return. For instance we can leverage the ERC positions so that the ERC strategy has the same mean rate of return. This can be achieved using a leverage ratio of 1.247. In this way we obtain a strategy with the same annualized mean return of 8.71% as the EW strategy and an annualized volatility of 8.47%, lower than the one of EW which is 10.52%. This reflects the fact that ERC has higher Sharpe ratio. Similarly, we can leverage the ERC positions such that the volatility is the same. The required leverage for this is 1.5497. Using this leverage the annualized mean return of ERC is 10.93%, which is better than EW, reflecting the higher Sharpe ratio of ERC.



4.2.2 CH Risk Parity Strategies

We do the same analysis for the swiss market.

	MV	ERC	$1/n$
μ	0.36%	0.41%	0.52%
μ_{ann}	4.41%	5.06%	6.40%
μ^c	3.37	3.85	4.85
σ	0.87%	1.12%	2.47%
$VaR_{0.01}$	0.10%	0.19%	1.34%
$ES_{0.01}$	0.16%	0.27%	2.11%
σ_{ann}	3.01%	3.88%	8.56%
VaR_{ann}	1.63%	3.02%	21.17%
ES_{ann}	2.55%	4.27%	33.37%
mean DD	0.51%	0.52%	9.51%
MDD	2.17%	4.01%	64.38%
S_σ	1.47	1.30	0.75
S_{VaR}	0.57	0.60	0.30
S_{ES}	0.46	0.45	0.24
S_{DD}	0.59	0.50	0.11
S_{MDD}	3.27%	2.98%	1.60%
ann. mean TO	8.99%	15.45%	0

In this case the strategy obtaining the best performance ratio is the MV strategy, with ERC being pretty close. The performance of EW is much lower. The drawdowns of the 50/50 strategy are really dramatic comparing with those of MV and ERC. The volatility, drawdowns and return paths of the threes strategies are with respect to each other similar as the ones of the USA market. The main difference between the american and swiss market over this period is the performance of the bonds markets. Swiss bonds have performed a lot better than USA ones. This is also the reason why MV portfolio performs so well here, minimum variance portfolios tend to be very concentrated in the instruments with lower volatility, in this case the bonds that have also a very high Sharpe ratio.

4.2.3 USA and CH Stocks & Bonds

We now want to consider mixed portfolios of USA/CH stocks and bonds. We ignore the risk given by the fluctuating exchange rate between USD and CHF.

Figure 4.12: Paths for MV, ERC and $1/n$ strategies of swiss stocks and bonds.

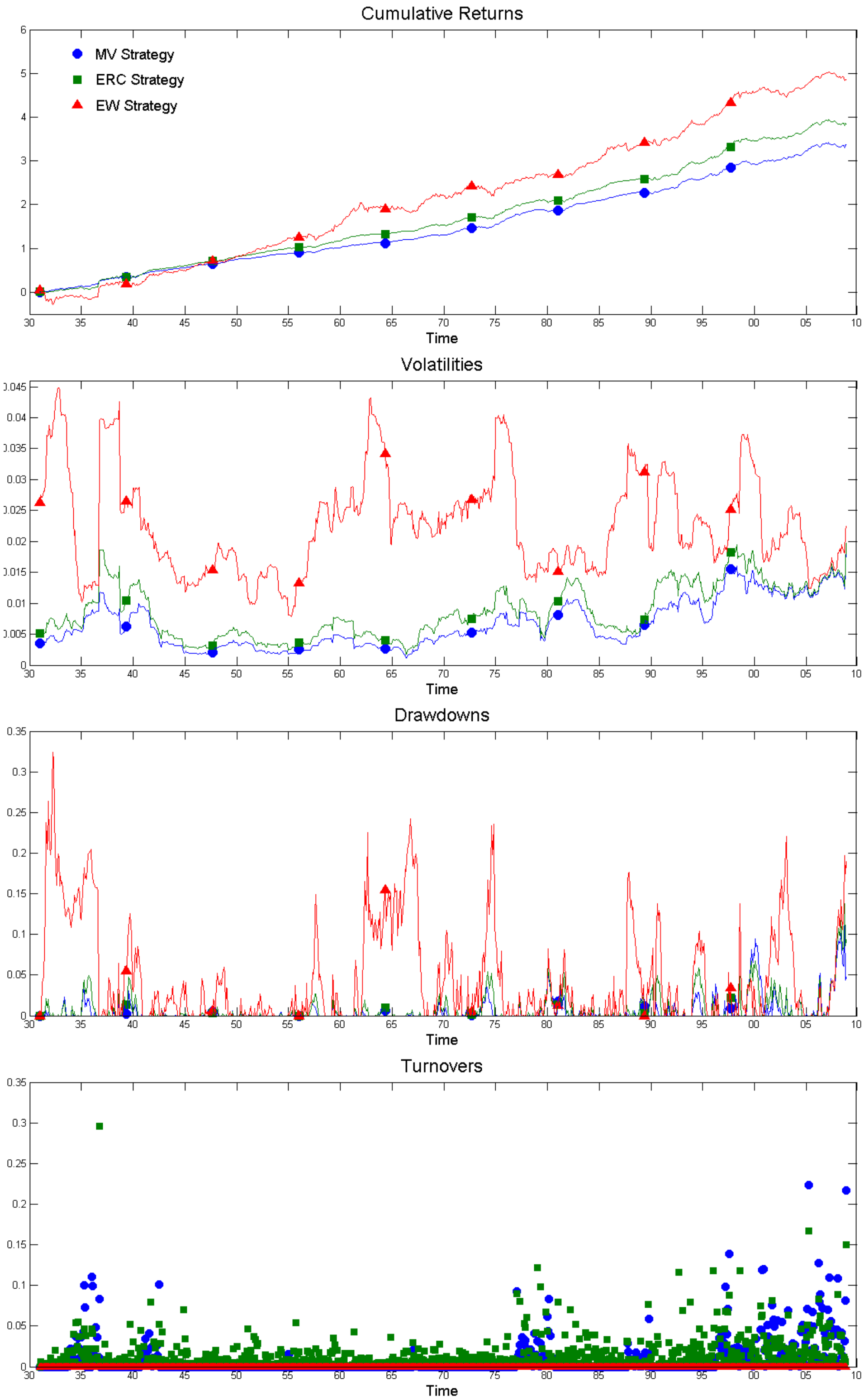
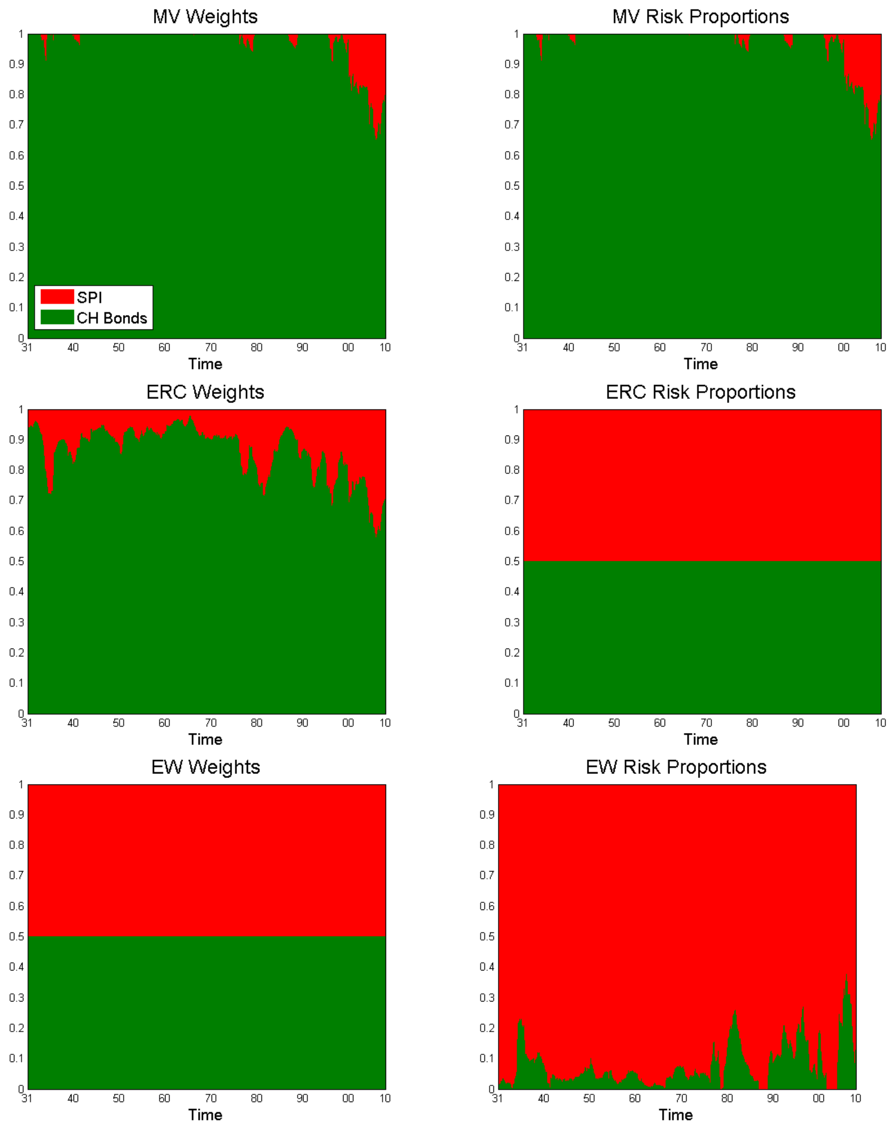


Figure 4.13: Swiss stocks/bonds weights and risk proportions.



	MV	ERC	$1/n$
μ	0.36%	0.45%	0.61%
μ_{ann}	4.46%	5.47%	7.55%
μ^c	3.41	4.16	5.69
σ	0.84%	1.09%	2.34%
$VaR_{0.01}$	1.66%	2.35%	6.67%
$ES_{0.01}$	2.64%	3.18%	8.29%
σ_{ann}	2.92%	3.79%	8.11%
VaR_{ann}	5.75%	8.13%	23.12%
ES_{ann}	9.16%	11.02%	28.71%
mean DD	0.51%	0.67%	2.69%
MDD	10.28%	11.94%	44.26%
S_σ	1.53	1.44	0.93
S_{VaR}	0.78	0.67	0.33
S_{ES}	0.49	0.50	0.26
S_{DD}	0.72	0.67	0.23
S_{MDD}	3.54%	3.73%	1.37%
ann. mean TO	61%	41%	0

As we can see from Figures 4.14 and 4.15, we obtain similar results as in the previous case.

Figure 4.14: Paths for MV, ERC and $1/n$ strategies considering the 4 assets.

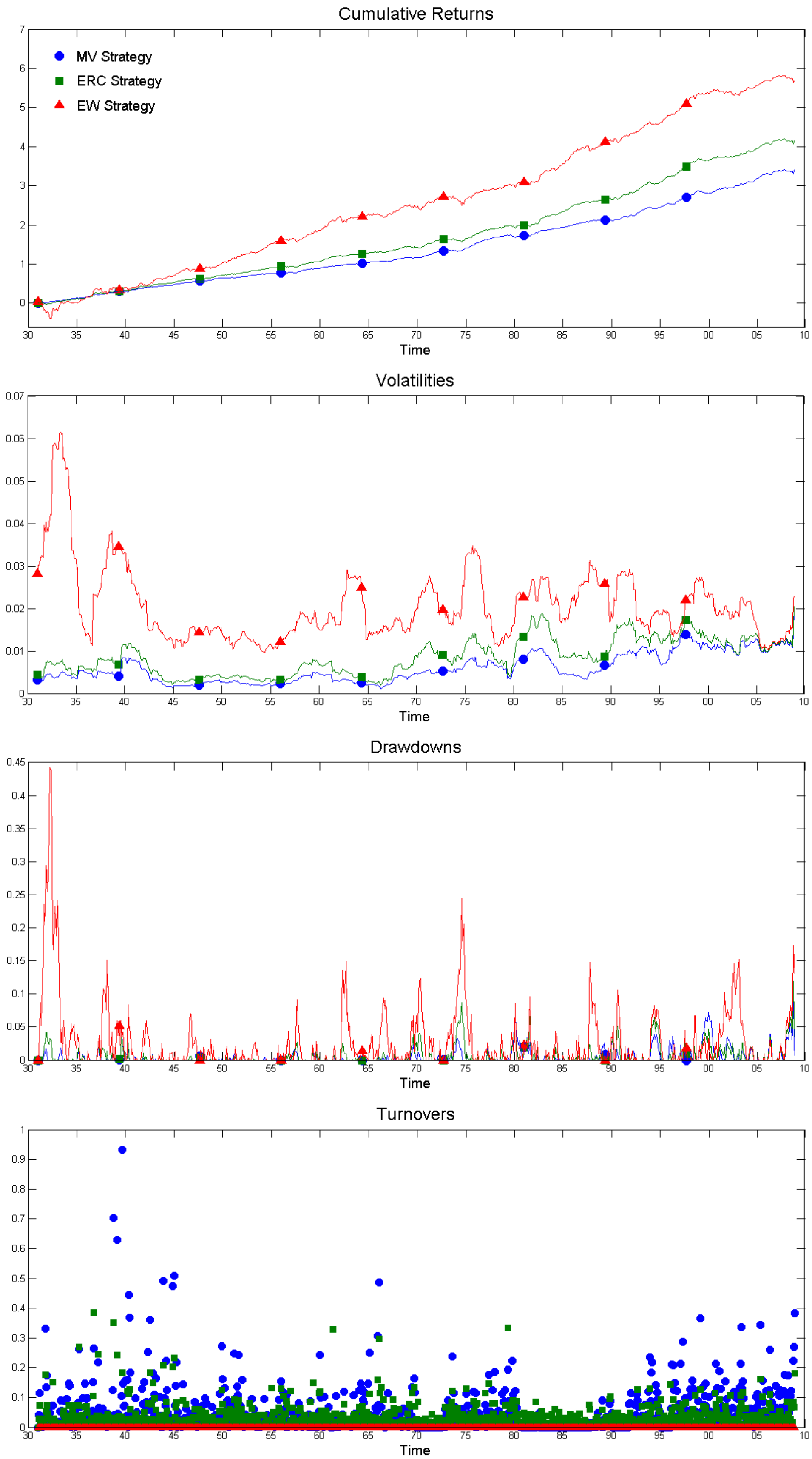
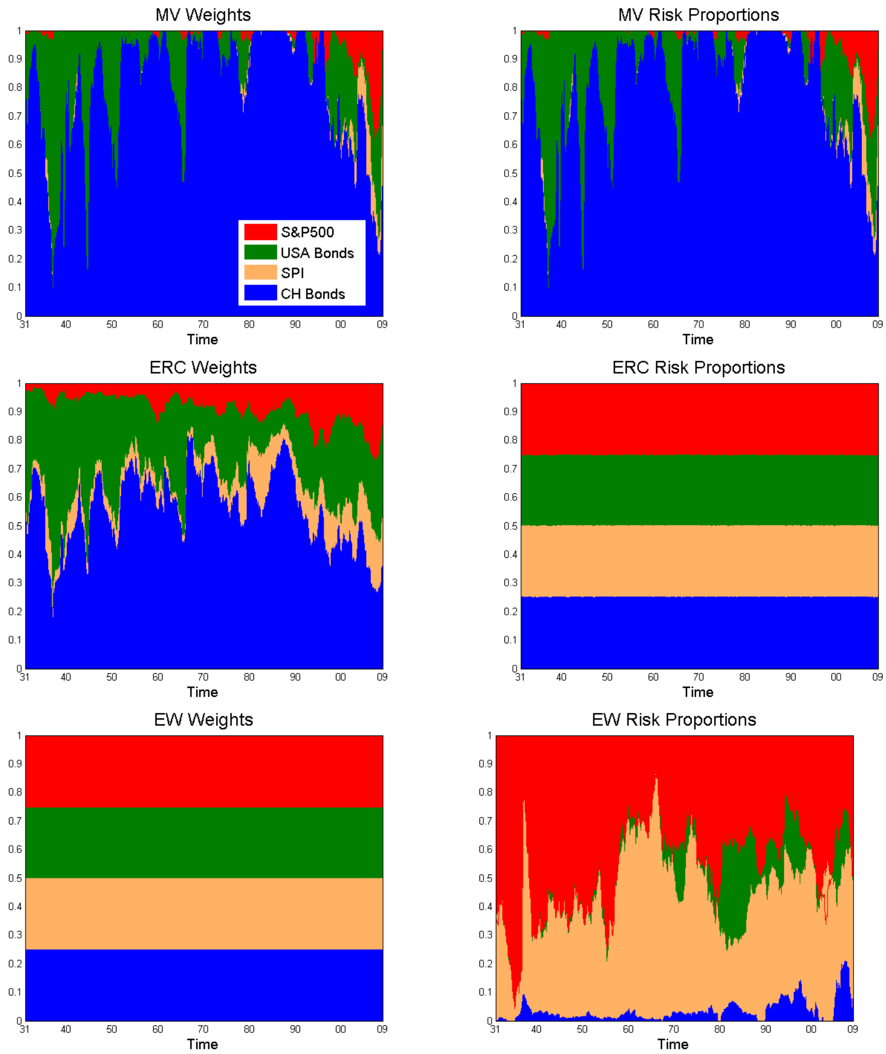


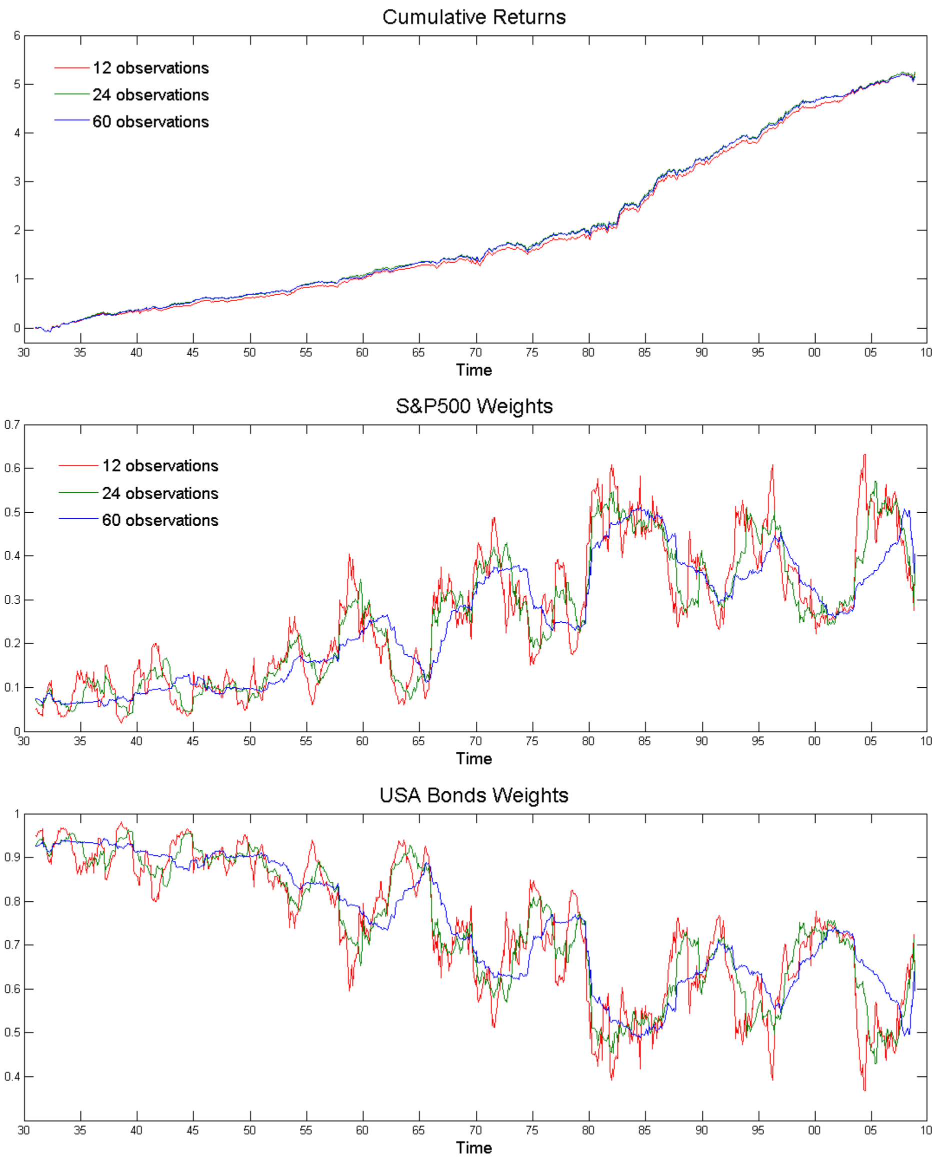
Figure 4.15: Weights and risk contributions of the 4 assets.



4.2.4 Stability Considerations

We conclude this example comparing the results for USA ERC strategies using different observation periods for the estimation of the covariance matrix. We consider 12, 24 and 60 observations.

	12	24	60
μ	0.56%	0.56%	0.55%
μ_{ann}	6.89%	6.93%	6.82%
μ^c	5.21	5.24	5.15
σ	1,97%	1.96%	1.98%
$VaR_{0.01}$	4.4%	4.56%	4.39%
$ES_{0.01}$	5.87%	5.99%	6.25%
σ_{ann}	6.81%	6.79%	6.87%
VaR_{ann}	15.25%	15.79%	15.19%
ES_{ann}	20.34%	20.74%	21.65%
mean DD	1.45%	1.39%	1.42%
MDD	15.62%	15.17%	18.76%
S_σ	1.01	1.02	0.99
S_{VaR}	0.45	0.44	0.45
S_{ES}	0.34	0.33	0.32
S_{DD}	0.38	0.40	0.39
S_{MDD}	3.56%	3.7%	2.94%
ann. mean TO	41.38%	23.35%	9.17%



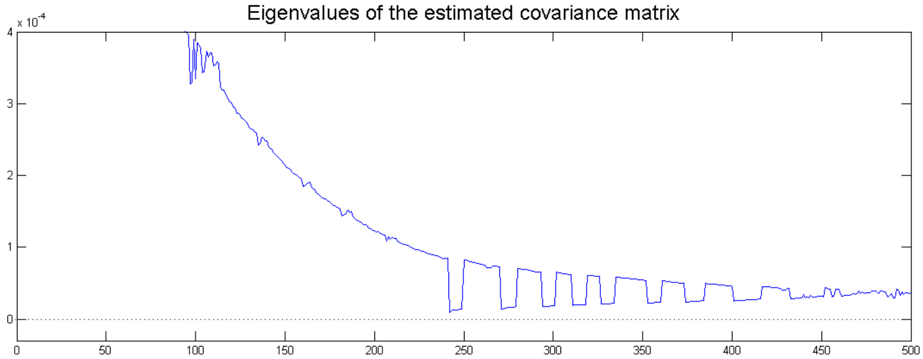
Reducing the number of observations produce more unstable weights because price changes have more impact in the covariances estimation if we use small samples. The annualized mean turnover is the only significant difference between the three. This means that by reducing the number of observations we get more sensitive strategies to price changes and we need to trade more. The performance statistics however do not change substantially. The return paths are also almost identical. We conclude that ERC strategies remain essentially the same in terms of performance statistics by reducing the number of observations. In terms of portfolio weights, the partition between bonds and stocks remain also quite similar, although by reducing the number of observations we observe more fluctuations in the portfolio weights which result in higher turnover. Reducing the number of observations from 60 to 24 and 12, the turnover increases from 9% to 23% and 41%, which is a strong increase in the transactions that makes the strategy less attractive in practice.

4.3 Large USA Stocks Portfolios

In this section we want to construct stocks portfolios using the 500 stocks of the U.S. based companies included in the S&P 500 index. We consider daily return observations and we rebalance the portfolio daily. In order to represent our results we divide the 500 stocks into 10 business categories. We first consider a single trading period from 29.04.2010 to 30.04.2010. We estimate the covariance matrix of the stock returns using the previous 250 observations, approximately one year of daily observations. In this example it is very important that we use the shrinkage estimator of Section 3.1 for the covariance matrix since we are in a small sample situation where we consider 500 instruments and only 250 observations. Sample covariances estimators in this case will always lead to a singular covariance matrix. We bring the estimated correlation matrix grouping the 500 stocks in the 10 categories.

Industrials	1									
Basic Materials	.88	1								
Financials	.83	.79	1							
Technology	.88	.84	.76	1						
Consumer Services	.93	.80	.80	.82	1					
Oil & Gas	.84	.90	.74	.79	.76	1				
Consumer Goods	.92	.81	.79	.82	.90	.78	1			
Utilities	.78	.71	.64	.68	.72	.74	.75	1		
Health Care	.73	.65	.62	.63	.69	.61	.74	.67	1	
Telecom	.69	.67	.63	.67	.64	.67	.68	.63	.62	1

We can observe that all the different categories are highly positively correlated. The eigenvalues of the estimated covariance matrix are all positive as desired. The following plot shows the smallest among the 500 eigenvalues.



For each of the 10 categories we give the total weight, the mean weight per stock, the maximum and the minimum weight in a single stock.

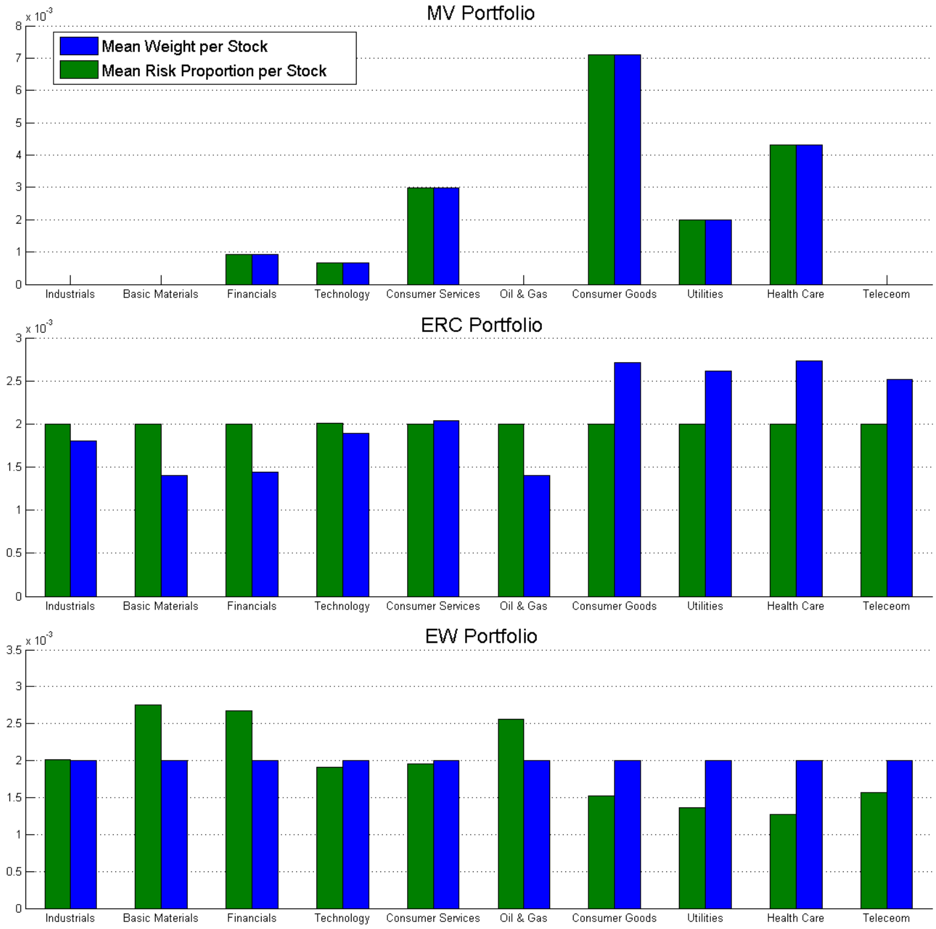
		MV			
	N. of Stocks	Weight	Mean	Max.	Min.
Industrials	72	0[1]	0	0	0
Basic Materials	26	0	0	0	0
Financials	82	7.564%	0.092%	6.2%	0
Technology	62	4.175%	4.75%	0	0
Consumer Services	75	22.286%	0.297%	16.186%	0
Oil & Gas	36	0	0	0	0
Consumer Goods	54	38.406%	0.71%	11.927%	0
Utilities	37	7.359%	0.199%	7.359%	0
Health Care	47	20.210%	0.430%	16.301%	0
Telecom	9	0	0	0	0

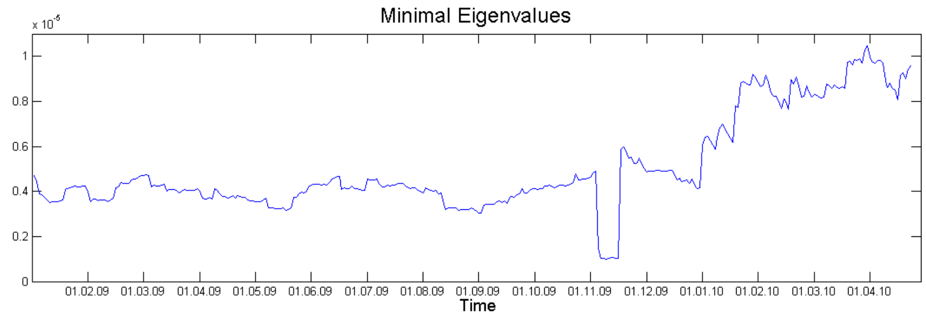
[1] Weights marked as 0 are not exactly zero, but very small values in the order of 10^{-10}

	ERC			
	Weight	Mean	Max.	Min.
Industrials	13.007%	0.181%	0.298%	0.093%
Basic Materials	3.649%	0.140%	0.261%	0.079%
Financials	11.858%	0.145%	0.337%	0.069%
Technology	11.723%	0.189%	0.390%	0.097%
Consumer Services	15.284%	0.204%	0.552%	0.071%
Oil & Gas	5.051%	0.140%	0.261%	0.094%
Consumer Goods	14.653%	0.271%	0.547%	0.093%
Utilities	9.654%	0.261%	0.432%	0.109%
Health Care	12.853%	0.273%	0.486%	0.127%
Telecom	2.269%	0.252%	0.386%	0.117%
	EW			
	Weight	Mean	Max.	Min.
Industrials	14.400%	0.200%	0.200%	0.200%
Basic Materials	5.200%	0.200%	0.200%	0.200%
Financials	16.400%	0.200%	0.200%	0.200%
Technology	12.400%	0.200%	0.200%	0.200%
Consumer Services	15.000%	0.200%	0.200%	0.200%
Oil & Gas	7.200%	0.200%	0.200%	0.200%
Consumer Goods	10.800%	0.200%	0.200%	0.200%
Utilities	7.400%	0.200%	0.200%	0.200%
Health Care	9.400%	0.200%	0.200%	0.200%
Telecom	1.800%	0.200%	0.200%	0.200%

MV portfolios are typically very concentrated and this can be observed very clearly in this example. Using the MV strategy we invest in only 5 categories and looking at maximal weights in each category we can deduce that more then 50% of our wealth is invested in only 5 stocks out of 500. In this example the ERC portfolio is much more homogeneous than in the two previous examples. This is because we are considering instruments that contribute to the total portfolio risk similarly with respect to each others.
We can observe that the optimization can be performed very accurately also in this case where we minimize a function of 500 variables. However, the optimization using the Matlab function *fmincon* takes quite a lot of time for this problem. For example, to obtain accurate equalization of risk contributions in the above one period example my Matlab implementation needed approximately 140 seconds on my machine and this means that if we want to backtest the strategy for the previous 10 years we will need approximately 4 days. For this reason I am going to consider only the period from 01.01.2009 to 29.04.2010 to compare the ERC strategy with the equally weighted strategy. For every trading period we compute the eigenvalues of the estimated covariance matrices, and in order to verify that all the eigenvalues are positive we plot the minimal eigenvalue for each period.

Figure 4.16: Weights and risk contributions at 29.04.2010.

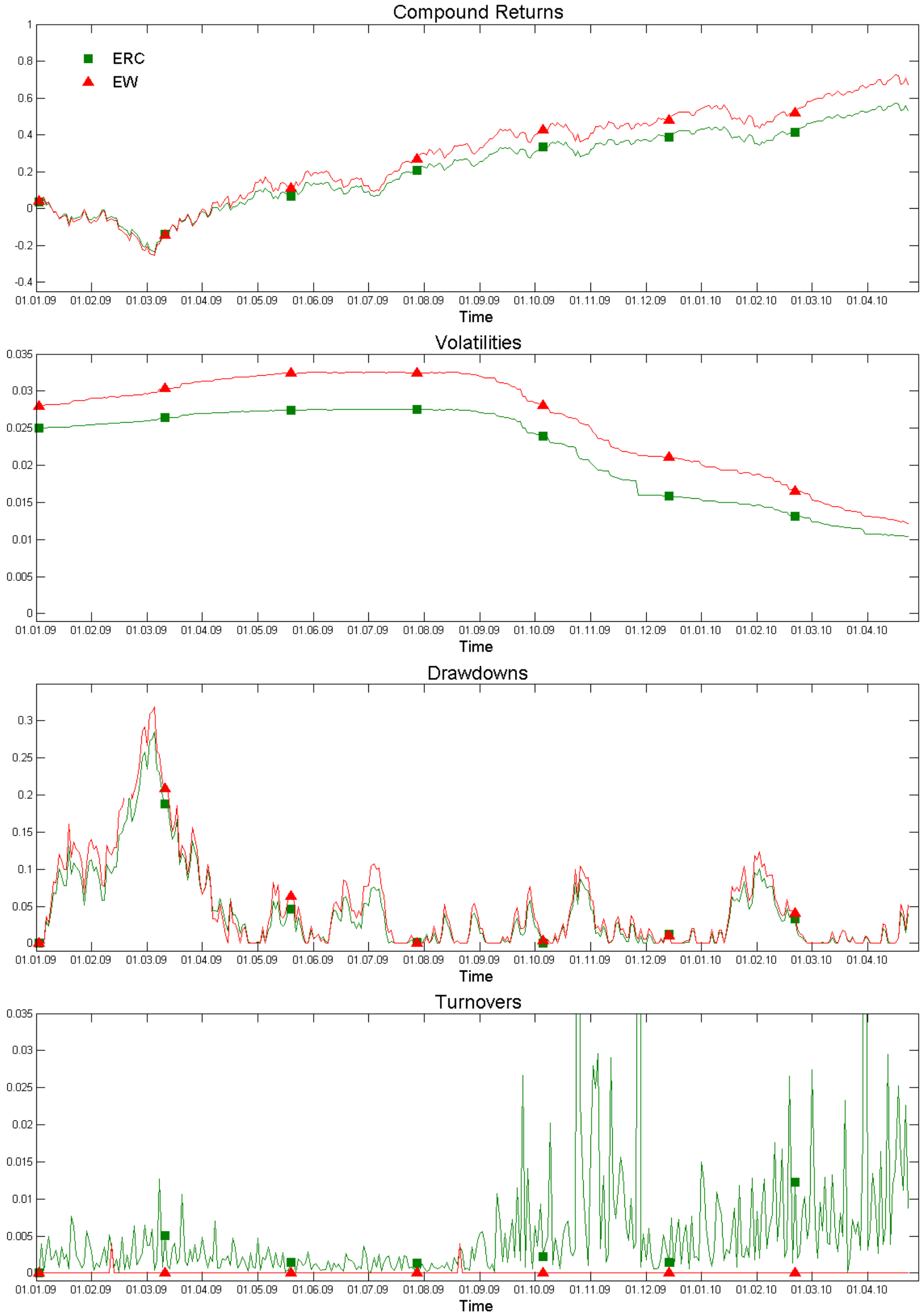




The portfolios statistics over this short period are similar comparing the two strategies (except for the turnover). This is because the set of instruments is very homogeneous.

	ERC	$1/n$
μ	0.14%	0.17%
μ_{ann}	40.35%	51.53%
μ^c	53.29%	67.37%
σ	1.56%	1.87%
$VaR_{0.01}$	4.68%	5.43%
$ES_{0.01}$	5.13%	5.96%
σ_{ann}	24.68%	29.57%
VaR_{ann}	73.98%	85.83%
ES_{ann}	81.09%	94.22%
mean DD	4.14%	5.01%
MDD	28.46%	31.88%
S_σ	1.64	1.74
S_{VaR}	0.55	0.6
S_{ES}	0.50	0.55
S_{DD}	3.27%	3.32%
S_{MDD}	0.48%	0.52%
ann. mean TO	1.52	0.58%

Figure 4.17: Paths for the ERC and 1/n strategies.



Chapter 5

Diversification Analysis

From a qualitative point of view the concept of diversification is quite clear: a portfolio is well-diversified if shocks in the individual components do not heavily impact on the overall portfolio. However, as Meucci [19] has noticed, there is no broadly accepted precise and quantitative definition of diversification. In this thesis we have considered diversification in terms of risk contributions of the individual assets. We have shown in Chapter 1 that a suitable definition of risk contribution can be obtained considering partial derivatives of the risk measure with respect to the portfolio components. We then constructed portfolio equalizing these risk contributions. In this chapter we want to present following [19] some methodologies to quantify diversification and apply them to one of the numerical examples of Chapter 4.

5.1 Diversification Measures

Consider as in the previous chapters n investment opportunities and denote by $\mathbf{w} \in \mathbb{R}^n$ a vector of portfolio weights satisfying the budget constraint $\mathbf{1}'\mathbf{w} = l$. We consider the volatility $\rho(\mathbf{w}) = \sqrt{\mathbf{w}'\Omega\mathbf{w}}$ as the measure of portfolio risk. As in the previous chapters $\Omega \in \mathbb{R}^{n\times n}$ denotes the covariance matrix of the assets returns. There are many naive ways of defining diversification measures in terms of portfolio weights:

$$\begin{aligned} D_{Her} &= 1 - \mathbf{w}'\mathbf{w}, \\ D_{BP} &= -\sum_{i=1}^{n} w_i \log(w_i), \\ D_{HK}^{\gamma} &= -\left(\sum_{i=1}^{n} w_i^{\gamma}\right)^{\frac{1}{\gamma-1}}, \gamma > 0. \end{aligned} \tag{5.1.1}$$

The first is known as *Herfindal index.* It is easy to see that it is zero for portfolios that are fully concentrated in one instrument and it achieves its maximum value $1 - \frac{l}{n}$ for the EW portfolio. The measure D_{BP} has been proposed by Bera and Park [3] in the case of long-only strategies, i.e. $w_i \geq 0$ for $i = 1, \ldots, n$. In this case, the weights can be interpreted as probability masses and diversification may be measured in terms of entropy. Simple computations show that D_{BP} achieves again its minimum zero for fully concentrated portfolios and its maximum $\log(n)$ for the EW portfolio. D^{γ}_{HK} has been proposed by Hannah and Kay [9]. It is easy to verify that $D^2_{HK} = D_{Her} - 1$ and

$$\begin{aligned}
\lim_{\gamma \to 1} D^{\gamma}_{HK}(\mathbf{w}) &= \lim_{\gamma \to 1} \left(- e^{\frac{1}{\gamma - 1} \log \left(\sum_{i=1}^{n} w_i^{\gamma} \right)} \right) = -e^{\lim_{\gamma \to 1} \frac{1}{\gamma - 1} \log \left(\sum_{i=1}^{n} w_i^{\gamma} \right)} \\
&= -e^{\lim_{\gamma \to 1} \frac{1}{\sum_{i=1}^{n} w_i^{\gamma}} \sum_{i=1}^{n} w_i^{\gamma} \log w_i} = -e^{\sum_{i=1}^{n} w_i \log w_i} \\
&= -e^{-D_{BP}(\mathbf{w})},
\end{aligned} \tag{5.1.2}$$

where the limit in the exponent has been computed using l'Hôpital's rule. These three quantities represent diversification only in terms of capital and do not take into account that different assets contribute differently to the total portfolio volatility. Equations (5.1.1) define suitable measures only if we are considering uncorrelated assets with equal volatilities. To take the risks into account practitioners often use the *differential diversification measure*,

$$D_{\text{diff}}(\mathbf{w}) = \sum_{i=1}^{n} w_i \sqrt{\Omega_{ii}} - \sqrt{\mathbf{w}' \Omega \mathbf{w}}. \tag{5.1.3}$$

By (1.2.5) the above expression is always greater or equal to zero. The first term is the sum of the stand-alone volatilities of the assets and the second term is the portfolio volatility. These two terms are the same when $\Omega = diag(\Omega_{11}, \ldots, \Omega_{nn})$. In general, the sum of the stand-alone risks exceeds the portfolio volatility. We interpret diversification as the effect responsible of this difference and define the diversification measure as the difference itself. The problem of this measure is that it does not highlight where the diversification comes from in a given portfolio. We now want to use the n risk contributions

$$RC_i(\mathbf{w}) = \frac{w_i (\Omega \mathbf{w})_i}{\sqrt{\mathbf{w}' \Omega \mathbf{w}}}, \quad i = 1, \ldots, n, \tag{5.1.4}$$

to construct a measure of diversification. We define

$$D_{RC}(\mathbf{w}) = \exp \left(- \left(\sum_{i,j=1}^{n} \Big(RC_i(\mathbf{w}) - RC_j(\mathbf{w}) \Big)^2 \right)^{\frac{1}{2}} \right). \tag{5.1.5}$$

D_{RC} achieves its maximum 1 for the ERC portfolio and becomes smaller if the differences between the risk contributions become larger. Another approach to quantify and manage diversification based on principal components has been proposed by Meucci [19]. In this article the author shows that it is always possible to find uncorrelated sources of risk which add up to the total portfolio risk. We consider the principal components decomposition of the positive semi-definite covariance matrix Ω,

$$\Omega = T\Lambda T', \tag{5.1.6}$$

where $T = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ and $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$. The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ denote the eigenvectors of Ω with eigenvalues $\lambda_1 \geq 0, \ldots, \lambda_n \geq 0$. Performing a transformation,

$$\tilde{\mathbf{w}} = T'\mathbf{w}, \tag{5.1.7}$$

of $\mathbf{w} \in \mathbb{R}^n$ in the principal basis, we obtain for the portfolio variance,

$$\rho^2(\mathbf{w}) = \mathbf{w}'\Omega\mathbf{w} = \mathbf{w}'T\Lambda T'\mathbf{w} = \tilde{\mathbf{w}}'\Lambda\tilde{\mathbf{w}} = \sum_{i=1}^{n} \tilde{w}_i^2 \lambda_i. \tag{5.1.8}$$

Dividing the last equation on both sides by $\sqrt{\mathbf{w}\Omega\mathbf{w}}$ we obtain

$$\sqrt{\mathbf{w}\Omega\mathbf{w}} = \sum_{i=1}^{n} \frac{\tilde{w}_i^2 \lambda_i}{\sqrt{\mathbf{w}\Omega\mathbf{w}}}, \tag{5.1.9}$$

which provides a decomposition of the portfolio volatility into volatility contributions along the coordinates of $\tilde{\mathbf{w}}$. This can be easily verified computing the partial derivatives of ρ with respect to the components of $\tilde{\mathbf{w}}$:

$$\frac{\partial \rho}{\partial \tilde{w}_i} = \frac{1}{2\sqrt{\sum_{i=1}^{n} \tilde{w}_i^2 \lambda_i}} 2\lambda_i \tilde{w}_i = \frac{\tilde{w}_i \lambda_i}{\sqrt{\mathbf{w}'\Omega\mathbf{w}}}. \tag{5.1.10}$$

Meucci [19] considers these additive components to define a measure of portfolio diversification. He defines the *diversification distribution*,

$$p_i(\mathbf{w}) = \frac{\tilde{w}_i^2 \lambda_i}{\mathbf{w}'\Omega\mathbf{w}} \in [0, 1], \tag{5.1.11}$$

for $i = 1, \ldots, n$. Using the normalized risk contributions along principal components p_i, $i = 1, \ldots, n$, instead of the usual risk contributions we have defined in Section 1.3, we have the advantage that these are always numbers between 0 and 1 (even if we allow for short position) that add up to 1. These define therefore probability masses and a portfolio is considered to be well-diversified if the diversification distribution is close to be uniform. To quantify the dispersion of this distribution the author has introduced a diversification measure based on its entropy,

$$D_{PC}(\mathbf{w}) = \exp\left(-\sum_{i=1}^{n} p_i(\mathbf{w}) \log\big(p_i(\mathbf{w})\big)\right). \tag{5.1.12}$$

This measure to which we refer as the diversification measure based on principal components achieves its minimum 1 for portfolios which are fully concentrated in one instrument and its maximum n if the diversification distribution is uniform. Furthermore, Meucci [19] suggests to combine this measure and expected returns to construct portfolios.

Definition 5.1.1. *Let $\alpha \in [0, 1]$, $\boldsymbol{\mu} \in \mathbb{R}^n$ be the vector of expected returns and $W \subset \mathbb{R}^n$ denote the set of admissible strategies. We define the mean-diversification efficient portfolio by*

$$\mathbf{w}^{MDE} = \arg\max_{\mathbf{w}\in W} \Big(\alpha\boldsymbol{\mu}'\mathbf{w} + (1-\alpha)D_{PC}(\mathbf{w})\Big). \tag{5.1.13}$$

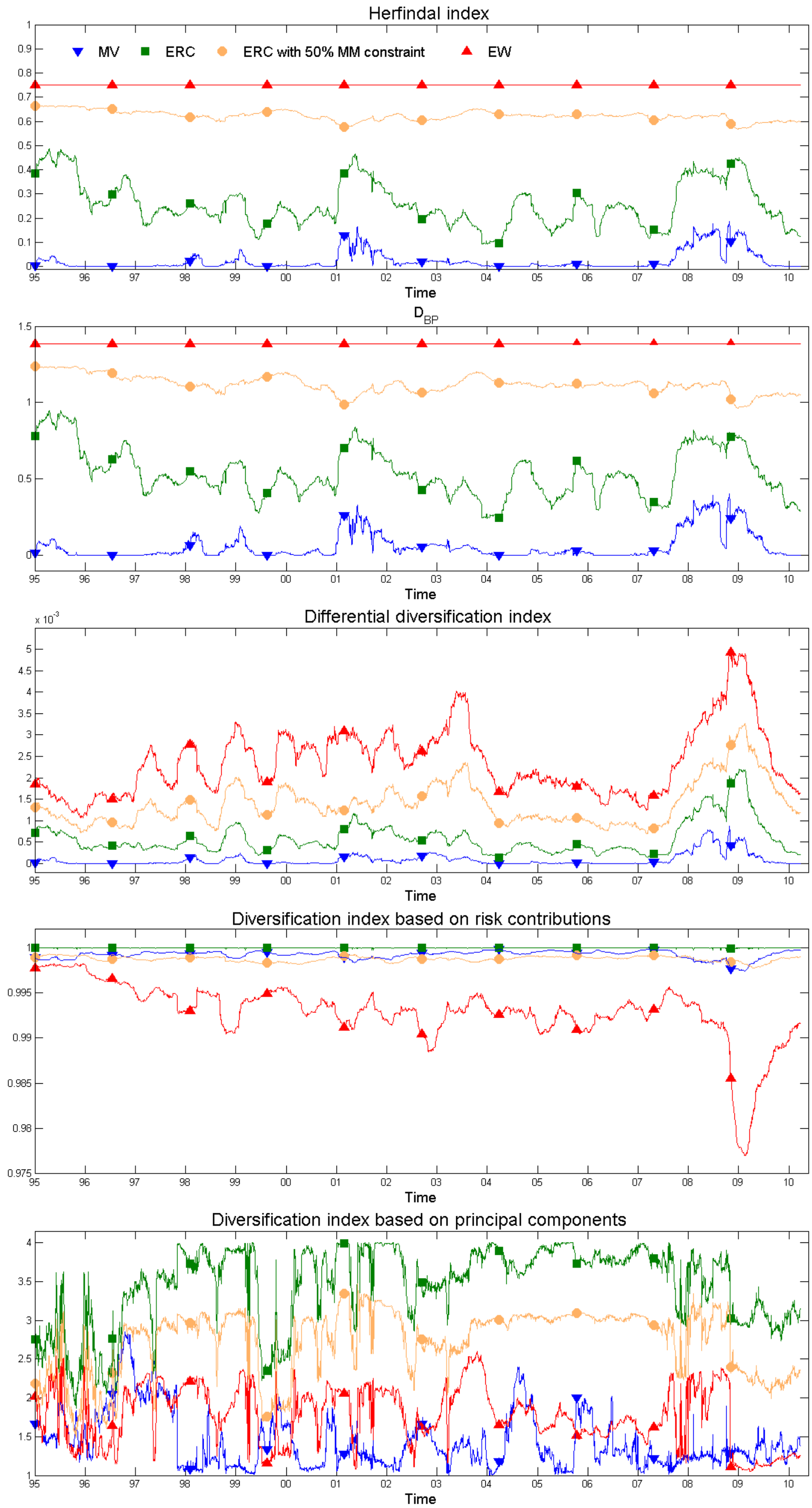
The parameter α defines the weight we assign to diversification or expected returns respectively in the optimization. We do not discuss such portfolios in this thesis.

5.2 Numerical Example

We consider the 4 assets example of Section 4.1. We compute daily the different diversification measures for the MV, ERC, ERC with weights constraints on money market and EW strategies discussing the results. The mean values of the diversification measures of the previous section are represented in the following table.

	MV	ERC	$w_3 \leq 0.7$	$w_3 \leq 0.6$	$w_3 \leq 0.5$	$1/n$
D_{Her}	0.0236	0.259	0.4649	0.5593	0.6236	0.75
D_{BP}	0.0601	0.5381	0.8687	1.0149	1.1196	1.3863
D_{Diff}	0.0001	0.0006	0.001	0.0012	0.0014	0.024
D_{RC}	0.9992	1	0.9994	0.9991	0.9987	0.9929
D_{PC}	1.4304	3.4545	2.9737	2.8179	2.7234	1.7661

We also graph in the following figure the paths of the diversification measures through time. We know from Section 4.1 that the MV strategy is the most concentrated one in terms of portfolio weights (almost everything in Money Market) and the EW is the most concentrated one in terms of risk contributions (almost all the risk in commodities and stocks). The ERC strategy is less concentrated in MM than MV. We can observe that the naive measures D_{Her}, D_{BP} and D_{diff} increase as the portfolios becomes more uniformly distributed in terms of weights. D_{Her} and D_{BP} achieve the maxima $3/4$ and $\log(4)$ respectively for the EW portfolio. The last two measures take into account risks and increase if the risk contributions are


Herfindal index
MV
ERC
ERC with 50% MM constraint
EW
Time
D_BP
Time
x 10^-3
Differential diversification index
Time
Diversification index based on risk contributions
Time
Diversification index based on principal components
Time

equalized. This was already clear from the definition for the measure D_{RC}. It is however interesting to observe that the same holds for the measure D_{PC} which is not based on the risk contributions with respect to which we are optimizing in the ERC approach. The maximum value of this measure for this example is 4 and this value is sometimes reached by the ERC optimal portfolio. This means that in some cases the optimal portfolio with respect to D_{RC} coincides with the one with respect to D_{PC}.

Conclusion

In this thesis we have discussed in detail how risk parity portfolios are constructed and we have performed some valuations based on historical data. In conclusion, we would like to explain the limitations of this approach that become evident from the numerical results of Chapter 4. We also want to bring in some ideas for further developments of this material.
We have already indicated in Chapter 2 that nothing prevents us from applying this methodology using other risk measures than the standard deviation (for example Value-at-Risk or Expected Shortfall). In practice however, using other risk measures than the volatility becomes difficult.
In Section 4.1 we have first considered this issue using two elliptical models for the asset returns (normally and t-distributed returns). In this case, we have analytical expressions for the risk contributions in terms of the asset returns covariances and expected values. This fact allowed us to easily compute equal contributions to risk portfolios using a shrinkage estimator for the covariance matrix (see Chapter 3). No significant difference in the allocation and performance has been observed in that example with respect to the allocation obtained using the volatility. Elliptical distributions do not always provide a good model for asset returns, especially in the case of instruments that exhibit high skewness. Generally, estimating the expected returns using the sample estimator leads to more unstable strategies. We have restricted ourselves to two special cases of elliptical models, but it would be also interesting to start with a completely general elliptical model and try to fit the parameters. The analytical expressions for the risk contributions in a general elliptical setting are known and given in Theorem 1.4.3. The quantities $\boldsymbol{\mu}$, Ω and $\rho(Y_1)$ need to be estimated. Some estimation methods are available for the expected returns $\boldsymbol{\mu}$ and the dispersion matrix Ω, see for example [17] Section 3.2.4. The quantity $\rho(Y_1)$ depends on the underlying spherical distribution. Empirical data for $\mathbf{Y}$ may be obtained from data for $\mathbf{X}$ computing $\Omega^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$, where $\Omega^{1/2}$ is the Cholesky factor of Ω. Estimating risk contributions in a general elliptical framework may lead to more accurate results than using a multivariate t-distribution as model for the asset returns.
In Chapter 3 we have also presented some classical estimators of VaR and ES risk contributions using empirical distributions. These approximations

are however quite rough and, as we have observed in Section 4.1.3, they lead to unstable portfolios (especially for VaR) with a lower performance than the other ones.
For these reasons and from our numerical experiments we conclude that the ERC approach can most easily and efficiently be implemented for the volatility. First of all, no particular distribution assumption is needed. Secondly, we are only required to estimate the variances and covariances for which a solid estimator has been discussed in Chapter 3. Moreover, as we have verified in Section 4.2, the portfolio weights and the performances are stable with respect to the covariance matrix estimation. From our numerical results it is also clear that the minimum variance portfolio is in most circumstances not applicable because of its extremely concentrated exposures (see for example Section 4.3), and that ERC portfolios are an efficient trade-off between $1/n$ and MV strategies.
Despite these nice properties of equal volatility contributions, it is important to realize that volatility is not the same as risk and it is not the only thing we need to know about the portfolio. Levels of volatilities that might appear under control may result in high losses, which may force the investors to cut down their exposures. For example, if we consider the 4 assets example of Section 4.1, we observe that the annualized volatility of the ERC strategy is 1.12%. In order to magnify the returns you may leverage the portfolio up to 8.16%, obtaining the same volatility as the equally weighted portfolio. This would lead to a maximum drawdown for the strategy of 29.22% in the period 2008-2009. The observed maximum drawdown in the same period for the $1/n$ portfolio is 64.38%. It is therefore important to realize that the portfolio volatility is not telling us everything about the potential losses, but other risk indicators need to be considered, especially when investors are amplifying returns using leverage. The use of leverage allows investors to reach higher returns, but at the same time leveraged investors may not be able to wait for prices to reach reasonable levels. Another problematic issue about leveraging positions is that the risks the investors are leveraging may not actually have positive return associated with them. Stocks and bonds allow the issuers to acquire the capital they need to function, and for this capital the issuers promise to the investors a positive return by taking on volatility. The situation looks different for commodities futures, for these types of contracts we have a buyer and a seller, and at the expiry date the contract will cease to exist. Why should the buyer be entitled to a systematically positive return and why should the commodity futures seller be prepared to accept a systematically negative return? The profits and losses are symmetrical to the buyer and the seller since both of them are taking on volatility, and where there is a winner, there is always also a loser. Asset classes without positive risk premiums associated with them are not suitable for long-term returns and leveraged positions in these assets should not be taken.

The use of volatility in Chapter 4 is also questionable since some of the asset classes under consideration have high negative skewness. See for example bonds and commodities futures in Section 4.1. Negative skewness means that the negative returns tend to be larger in magnitude than the positive ones. In this case the volatility is not able to capture what the true downside risk is. This is particularly pronounced in the context of credit risks. The borrower will either pay back the loan or default, in case he pays back the loan, the return is slightly positive, in case he doesn't pay back the loan the return is strongly negative. It is impossible to quantify precisely the risk in this case for a given period of observations by simply estimating volatility contributions. Combining the unknown risks of the investment with leverage can be very dangerous.

In Section 4.3 we have constructed risk parity portfolios for assets of the same risk class (500 stocks). In this situation where the risk is very homogeneously distributed among the assets there is no significant difference between balancing risks and splitting the wealth equally among the assets. The results are represented in Figures 4.15 and 4.16. The performance remains very similar (even lower) and the ERC approach requires a much higher turnover than the equally weighted allocation. We conclude therefore that in such a situation it is not worth to switch from $1/n$ to risk balanced allocations.

There are of course many improvement that can be done to the topics discussed in this thesis. In Chapter 3 we limited ourselves only to the estimation of risk contributions using historical data. Methods based on Monte-Carlo simulations may also be developed. In order to apply the ERC principle to more appropriate risk measures such as Expected Shortfall or Semi-Variance, more sophisticated techniques to estimate the risk contributions need to be discussed. Some more accurate techniques for Value-at-Risk and Expected Shortfall can be found in [20].

In Chapter 5, we have introduced a different notion of risk contributions based on the principal components of the covariance matrix and we have seen that these contributions correspond exactly to the partial derivatives of the volatility along the principal components. We haven't tried to compute the optimal portfolio maximizing the measure D_{PC} instead of D_{RC}. In Section 5.2, we have observed computing D_{PC} for the portfolios of Section 4.1 that ERC portfolios tend to have high diversification values in terms of D_{PC} as well and that in some cases they exhibit values that are pretty close to the maximum. This happens when the covariance matrix is very close to be diagonal, but in general the two allocations differ. It would be interesting to compare the allocation and the performance of the two principles in some examples.

In Section 4.1 we have considered weights constraints in the Money Market and observed higher returns without affecting significantly the performance ratios (for some constraints we have even observed slightly higher

performance ratios). This motivates the following question: how can we deviate from the ERC diversified portfolio in order to increase performance? Meucci [19] has proposed the *mean-diversification efficient portfolio* (Definition 5.1.1). Using this selection criterion we deviate from the perfectly diversified portfolio (diversification measured using principal components as explained in Section 5.1) incorporating expected returns information.
From our numerical results and considering the above remarks we conclude that the ERC approach for the volatility as risk measure provides an interesting benchmark, based only on risk diversification, that has some limitations but also some advantages: it is based on a quite simple mathematical framework, it can be easily implemented having an efficient optimization solver (such as in Matlab), it provides higher performance ratios than the traditional equally weighted allocation, and it is based on a quite intuitive principle that can also be explained effortless to people without a mathematical background in portfolio construction.

Appendix A

Spherical and Elliptical Distributions

We introduce formally and present the basic properties of this important class of multivariate distributions. We begin with the special case of spherical distributed random variables.

Definition A.1. *A random vector* $\mathbf{R} = (R_1, \ldots, R_n)$ *has spherical distribution if, for every orthogonal map* $U \in \mathbb{R}^{n \times n}$ *(i.e. a map satisfying* $U'U = UU' = I_n$*) it holds*

$$U\mathbf{R} \overset{(d)}{\sim} \mathbf{R}, \tag{A.1}$$

i.e. spherically distributed random variables are those with invariant distribution under orthogonal transformations.

The following theorem provides the main ways of characterizing spherical distributions.

Theorem A.1. *The following statements are equivalent.*

1. $\mathbf{R}$ *is spherically distributed.*
2. *The characteristic function of* $\mathbf{R}$ *satisfies, for any* $\mathbf{t} \in \mathbb{R}^n$, $\phi_{\mathbf{R}}(\mathbf{t}) = E[e^{i\mathbf{t}'\mathbf{R}}] = \psi(\mathbf{t}'\mathbf{t})$ *for some scalar function* ψ.
3. *For every* $\mathbf{a} \in \mathbb{R}^n$, *it holds* $\mathbf{a}'\mathbf{R} \overset{(d)}{\sim} |\mathbf{a}|R_1$, *where* $|\cdot|$ *denotes the Euclidean norm.*

Property 2 tells us that the distribution of a spherical random variable is completely described by a scalar function ψ. We introduce therefore the notation $\mathbf{R} \sim S_n(\psi)$, for spherical distributed random variables. Property 3 is very important and has been used in Theorem 1.2.1. It allows to characterize linear combinations of spherical random vectors. The proof of this result can be found in [17] Theorem 3.19. It can be easily verified using [17] Theorem 3.22 that, for $\mathbf{R} \sim S(\psi)$, $E[\mathbf{R}] = 0$ and that the covariance matrix (if the second moments are finite) is given by $\Omega = c(\psi)I_n$ where $c(\psi) > 0$ is a positive constant.

Example A.1. Consider a random vector $\mathbf{R} \in \mathcal{N}(\mathbf{0}, I_n)$ with the standard uncorrelated multinormal distribution. $\mathbf{R}$ is clearly spherical, since its characteristic function is given by

$$\phi_{\mathbf{R}}(t) = E[e^{i\mathbf{t}'\mathbf{R}}] = \psi(\mathbf{t}'\mathbf{t}) \tag{A.2}$$

with $\psi(x) = e^{-\frac{1}{2}x}$. We can see spherical distributions as an extension of standard multinormal distributions $\mathcal{N}(\mathbf{0}, I_n)$.The reason why we introduce this generalization is that these distributions allow to describe heavy tails.

In Example 1.1.4 we describe a model based on the t-distribution which is also a spherical distribution. Elliptical distributions are affine transformation of spherical distributions.

Definition A.2. $\mathbf{R}$ *has an elliptical distribution if*

$$\mathbf{R} \overset{(d)}{\sim} A\mathbf{Y} + \boldsymbol{\mu} \tag{A.3}$$

where $\mathbf{Y} \in S_k(\psi)$*,* $A \in \mathbb{R}^{n\times k}$ *and* $\boldsymbol{\mu} \in \mathbb{R}^n$*.*

The characteristic function of an elliptical distributed random variable is given by

$$\phi_{\mathbf{R}}(\mathbf{t}) = E\big[e^{i\mathbf{t}'(A\mathbf{Y}+\boldsymbol{\mu})}\big] = e^{i\mathbf{t}'\boldsymbol{\mu}}E\big[e^{i(A'\mathbf{t})'\mathbf{Y}}\big] = e^{i\mathbf{t}'\boldsymbol{\mu}}\psi(\mathbf{t}'\Omega\mathbf{t}) \tag{A.4}$$

where $\Omega = A'A$. The distribution of an elliptical distributed random variable is therefore determined by the parameters $\boldsymbol{\mu}$, called the *location vector*, Ω, referred to as *dispersion matrix*, and the function ψ, the *characteristic generator* of the corresponding spherical distribution. We use the notation $\mathbf{R} \sim E_n(\boldsymbol{\mu}, \Omega, \psi)$. Note that Ω and ψ are not uniquely determined by the distribution, but only up to a positive constant (consider the pair $(c\Omega, \psi(\cdot/c))$ for $c > 0$). For elliptical distributed random variables we have $E[\mathbf{R}] = \boldsymbol{\mu}$ and, provided second moments exist, its covariance matrix is a multiple of Ω. It is therefore possible to find an elliptical representation $E_n(\boldsymbol{\mu}, \Omega, \psi)$ so that Ω is the covariance matrix of $\mathbf{R}$. However we do not impose this with the above notation. We now look at some basic properties of this class of random variables. If we take linear combinations of elliptical random vectors these remain elliptical: let $\mathbf{R} \sim E_n(\boldsymbol{\mu}, \Omega, \psi)$ and for any $B \in \mathbb{R}^{k\times n}$ and $\mathbf{b} \in \mathbb{R}^n$, we have

$$B\mathbf{R} + \mathbf{b} \sim E_n(B\boldsymbol{\mu} + \mathbf{b}, B'\Omega B, \psi). \tag{A.5}$$

Particularly important in our case is the fact that $\mathbf{m}'\mathbf{R} \sim E_n(\mathbf{m}'\boldsymbol{\mu}, \mathbf{m}'\Omega\mathbf{m}, \psi)$. From this it follows that marginal distributions of $\mathbf{R}$ must be also elliptical distributions: separate the coordinates of the random vector according to $\mathbf{R} = (\mathbf{R}^{(l)}, \mathbf{R}^{(n-l)})$, where $\mathbf{R}^{(l)}$ is the vector of the first l coordinates and $\mathbf{R}^{(n-k)})$ is the vector of the remaining coordinates, then we have

$$\mathbf{R}^{(l)} \sim E_l(\boldsymbol{\mu}^{(l)}, \Omega^{(l)}, \psi), \quad \mathbf{R}^{(n-l)} \sim E_{n-l}(\boldsymbol{\mu}^{(n-l)}, \Omega^{(n-l)}, \psi) \tag{A.6}$$

where $\boldsymbol{\mu} = (\boldsymbol{\mu}^{(l)}, \boldsymbol{\mu}^{(n-l)})$, $\Omega^{(l)} \in \mathbb{R}^{l \times l}$ and $\Omega^{(n-l)} \in \mathbb{R}^{(n-l)\times(n-l)}$ are the corresponding sub vectors/matrices.

Appendix B

Inequality Constraints and Kuhn-Tucker Conditions

We are interested in minimizing a function in several variables under some equality and possibly also inequality constraints. Formally the problem is the following,

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}) \quad \text{subject to} \quad \begin{cases} c_i(\mathbf{x}) = 0 & i \in \mathcal{E} \\ c_i(\mathbf{x}) \geq 0 & i \in \mathcal{I} \end{cases}, \tag{B.1}$$

where f and c_i are continuously differentiable functions on $\mathbb{R}^n$, and $\mathcal{E}$ and $\mathcal{I}$ are two finite set of indices that correspond to the equality and inequality constraints respectively. We define by $F = \{\mathbf{x} \in \mathbb{R}^n | c_i(\mathbf{x}) = 0 \ \text{ for all } \ i \in \mathcal{E}, c_i(\mathbf{x}) \geq 0 \ \text{ for all } \ i \in \mathcal{I}\}$ the set of all *feasible* vectors. We want to restrict ourselves to vectors satisfying certain constraints qualifications.

Definition B.1. *The active set $\mathcal{A}(\mathbf{x})$ at any feasible $\mathbf{x} \in F$ is defined as*

$$\mathcal{A}(\mathbf{x}) = \mathcal{E} \cup \{i \in \mathcal{I} : c_i(\mathbf{x}) = 0\}, \tag{B.2}$$

i.e. it consists of the equality constraints indices together with the indices of the inequality constraints where equality is reached.

Definition B.2. *Given a vector $\mathbf{x}$ and the corresponding active set $\mathcal{A}(x)$ we say that the linear independence constraint qualification, abbreviated LICQ, holds if the gradients*

$$\nabla c_i(\mathbf{x}) \quad \textit{for} \quad i \in \mathcal{A}(\mathbf{x}) \tag{B.3}$$

are linearly independent.

The following first order conditions are the foundation of the most popular algorithms to solve (B.1), in the literature people usually refer to them as Kuhn-Tucker conditions. A vector $\mathbf{x}^*$ is called a *local solution* of (B.1) if

$\mathbf{x}^* \in F$ and there is a neighborhood $\mathcal{N}$ of $\mathbf{x}^*$ such that $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x} \in \mathcal{N} \cap F$.

Theorem B.0.1. *Define the Lagrange function for the minimization problem (B.1) as*

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(\mathbf{x}). \tag{B.4}$$

Suppose that $\mathbf{x}^*$ *is a local solution of (B.1) and that the LICQ holds at* $\mathbf{x}^*$*. Then, there is a Lagrange multiplier vector* $\boldsymbol{\lambda}^*$*, with components* λ_i^**,* $i \in \mathcal{E} \cup \mathcal{I}$*, such that the following conditions are satisfied at* $(\mathbf{x}^*, \boldsymbol{\lambda}^*)$

$$\begin{cases} \nabla_{\mathbf{x}} L(\mathbf{x}^*, \boldsymbol{\lambda}^*) = \mathbf{0}, & \\ c_i(\mathbf{x}^*) = 0, & \textit{for all} \;\; i \in \mathcal{E}, \\ c_i(\mathbf{x}^*) \geq 0, & \textit{for all} \;\; i \in \mathcal{I}, \\ \boldsymbol{\lambda}^* \geq 0, & \textit{for all} \;\; i \in \mathcal{I}, \\ \boldsymbol{\lambda}^* c_i(\mathbf{x}^*) = 0, & \textit{for all} \;\; i \in \mathcal{E} \cup \mathcal{I}. \end{cases} \tag{B.5}$$

The proof of this theorem is quite complex and can be found in [21] Section 12.4.

Appendix C

Sequential Quadratic Programming

As we have seen in Chapter 2 in order to find optimal portfolio strategies we need to solve numerical minimization problems of the form (B.1). We now discuss one efficient method to do this which is implemented in Matlab under the function *fmincon* that minimizes general functions of several variables under general constraints. It is important to understand at least qualitatively how this function works since it is the source of all our numerical results presented in Chapter 4. It is important to note that there are three different algorithms used by this function, and the one we are going to use and discuss here is the one referred to as *active-set algorithm* in the description of the function available in the Matlab Product Help. This method tries to compute the Lagrange multipliers from the Kuhn-Tucker conditions (B.5) explicitly by solving in every iteration a quadratic programming subproblem. These types of methods are usually referred to in the literature as *sequential quadratic programming* methods, abbreviated SQP methods. We want to sketch the idea used by SQP algorithms, explaining very briefly how it is implemented in the Matlab function we have used in Chapter 4. Complete description about how to construct QP and SQP algorithms can be found in [21] Chapter 16 and Chapter 18. We begin considering only equality constraints. The method consists mainly in deriving an appropriate QP subproblem. This can found applying the Newton's method for non-linear equations to the Kuhn-Tucker conditions. The Lagrange function for a problem with m equality constraints is given by

$$L(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) - \boldsymbol{\lambda}'\mathbf{c}(\mathbf{x}), \tag{C.1}$$

where $\mathbf{x} \in \mathbb{R}^n$, $\boldsymbol{\lambda} \in \mathbb{R}^m$ and $\mathbf{c}(\mathbf{x}) = (c_1(\mathbf{x}), \dots, c_m(\mathbf{x}))$ denotes the equality constraints. Let $J_c(\mathbf{x})$, with

$$J_c(\mathbf{x})' = [\nabla c_1(\mathbf{x}), \dots, \nabla c_m(\mathbf{x})], \tag{C.2}$$

be the Jacobi matrix of the constraints. The Kuhn-Tucker conditions (B.5) for this problem consisting only of equality constraints are given by

$$F(\mathbf{x}, \boldsymbol{\lambda}) = \begin{pmatrix} \nabla f(\mathbf{x}) - J_c(\mathbf{x})'\boldsymbol{\lambda} \\ \mathbf{c}(\mathbf{x}) \end{pmatrix} = \mathbf{0}, \tag{C.3}$$

where $\mathbf{0} \in \mathbb{R}^{n+m}$ denotes the vector of zeros. Any solution of the problem for which J_c has full rank must satisfy the above equation. The Newton's iteration for the solution of (C.3) is given by,

$$\begin{pmatrix} \mathbf{x}_{k+1} \\ \boldsymbol{\lambda}_{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_k \\ \boldsymbol{\lambda}_k \end{pmatrix} + \begin{pmatrix} \mathbf{p}_k \\ \mathbf{p}_\lambda \end{pmatrix} \tag{C.4}$$

with $(\mathbf{p}_k, \mathbf{p}_\lambda)$ being the solution of

$$J_F(\mathbf{x}_k, \boldsymbol{\lambda}_k) \begin{pmatrix} \mathbf{p}_k \\ \mathbf{p}_\lambda \end{pmatrix} = \begin{pmatrix} -\nabla f(\mathbf{x}_k) + J_c(\mathbf{x}_k)'\boldsymbol{\lambda}_k \\ -\mathbf{c}(\mathbf{x}_k) \end{pmatrix}, \tag{C.5}$$

where

$$J_F(\mathbf{x}_k, \boldsymbol{\lambda}_k) = \begin{bmatrix} \nabla^2_{xx} L(\mathbf{x}_k, \boldsymbol{\lambda}_k) & -J_c(\mathbf{x}_k)' \\ J_c(\mathbf{x}_k) & 0 \end{bmatrix} \tag{C.6}$$

is the Jacobi matrix of F at the iterate $(\mathbf{x}_k, \boldsymbol{\lambda}_k)$. This iteration is well-defined and convergent if J_F is non singular and if the initial guess for the iteration is near enough to the optimum. This can be ensured making the following assumptions.

1. The matrix $J_c(\mathbf{x})$ has full row rank.
2. The matrix $\nabla^2_{xx} L(\mathbf{x}, \boldsymbol{\lambda})$ is positive definite on the tangent space of the constraints, i.e. $\mathbf{p}'\nabla^2_{xx} L(\mathbf{x}, \boldsymbol{\lambda})\mathbf{p} > 0$ for all $\mathbf{p} \neq \mathbf{0}$ with $J_c(\mathbf{x})\mathbf{p} = \mathbf{0}$.

Under these assumptions (see [21] Theorem 18.4) the above iteration converges quadratically and provides a very efficient algorithm to solve equally constrained problems if the initial guess is close to the solution. It is now interesting to observe that the same iteration can be obtained by solving instead of the initial problem the following QP problem.

$$\begin{aligned} \min_{\mathbf{p} \in \mathbb{R}^n} \quad & f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)'\mathbf{p} + \frac{1}{2}\mathbf{p}'\nabla^2_{xx} L(\mathbf{x}_k, \boldsymbol{\lambda}_k)\mathbf{p} \\ \text{subject to} \quad & J_c(\mathbf{x}_k)\mathbf{p} + \mathbf{c}(\mathbf{x}_k) = \mathbf{0}. \end{aligned} \tag{C.7}$$

Note that if the above assumptions are satisfied there exists a solution $(\mathbf{p}_k, \mathbf{l}_k)$ satisfying the corresponding Kuhn-Tucker conditions for the above QP problem.

$$\begin{cases} \nabla^2_{xx} L(\mathbf{x}_k, \boldsymbol{\lambda}_k)\mathbf{p}_k + \nabla f(\mathbf{x}_k) - J_c(\mathbf{x}_k)'\mathbf{l}_k = 0 \\ J_c(\mathbf{x}_k)\mathbf{p}_k + \mathbf{c}(\mathbf{x}_k) = \mathbf{0}. \end{cases} \tag{C.8}$$

This solution is closely related to the Newton's iteration. To see this, subtract $(J_c(\mathbf{x}_k)'\lambda_k, 0)$ from (C.5) to obtain

$$\begin{bmatrix} \nabla^2_{xx} L(\mathbf{x}_k, \boldsymbol{\lambda}_k) & -J_c(\mathbf{x}_k)' \\ J_c(\mathbf{x}_k) & 0 \end{bmatrix} \begin{pmatrix} \mathbf{p}_k \\ \boldsymbol{\lambda}_{k+1} \end{pmatrix} = \begin{pmatrix} -\nabla f(\mathbf{x}_k) \\ -c(\mathbf{x}_k) \end{pmatrix}. \tag{C.9}$$

From this and by non-singularity of the coefficient matrix it follows that $\boldsymbol{\lambda}_{k+1} = \mathbf{l}_k$ and that $\mathbf{p}_k$ is a solution of (C.5) and (C.7). Note that the equality constraints allow to replace $\nabla f(\mathbf{x}_k)'\mathbf{p}$ by $\nabla_x L(\mathbf{x}_k, \boldsymbol{\lambda}_k)'\mathbf{p}$ in (C.7), and the function to minimize becomes a quadratic approximation of L at $(\mathbf{x}_k, \boldsymbol{\lambda}_k)$. This motivates the choice of our QP problem: we replace the initial problem by the problem of minimizing L under the equality constraints, then we make a quadratic approximation of L and a linear approximation of the constraints and we solve the approximated problem. The same idea can be extended to the more generic problem with inequality constraints: for the iterate $(\mathbf{x}_k, \boldsymbol{\lambda}_k)$ we consider the QP

$$\begin{aligned} \min_{\mathbf{p} \in \mathbb{R}^n} \quad & f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)'\mathbf{p} + \frac{1}{2}\mathbf{p}'\nabla^2_{xx} L(\mathbf{x}_k, \boldsymbol{\lambda}_k)\mathbf{p} \\ \text{subject to} \quad & \nabla c_i(\mathbf{x}_k)'\mathbf{p} + c_i(\mathbf{x}_k) = \mathbf{0} \quad \text{for} \quad i \in \mathcal{E} \\ \text{and} \quad & \nabla c_i(\mathbf{x}_k)'\mathbf{p} + c_i(\mathbf{x}_k) \geq \mathbf{0} \quad \text{for} \quad i \in \mathcal{I}. \end{aligned} \tag{C.10}$$

The basic structure of the SQP Algorithm is then the following. We begin setting $k = 0$ and choosing an initial pair $(\mathbf{x}_0, \boldsymbol{\lambda}_0)$. We then repeat until a certain convergence criterion is satisfied the following steps: we evaluate $f(\mathbf{x}_k), \nabla f(\mathbf{x}_k), \nabla^2_{xx} L(\mathbf{x}_k, \boldsymbol{\lambda}_k), \mathbf{c}(\mathbf{x}_k)$ and $J_c(\mathbf{x_k})$, solve (C.10) to obtain $\mathbf{p}_k$ and the corresponding Lagrange multiplier $\mathbf{l}_k$ and then set $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{p}_k$ and $\boldsymbol{\lambda}_{k+1} = \mathbf{l}_k$. Different SQP algorithms are normally different in the way they solve the QP problem or in the way they treat the Hessian $\nabla^2_{xx} L$, but the basic structure is the one described above. In the Matlab implementation provided by the function *fmincon* this is implemented in the following way. At each iteration the QP subproblem is solved using an *active-set* method for convex quadratic programming. A complete description of these methods can be found in [21] Section 16.5. At each iteration an estimate of the Hessian $\nabla^2_{xx} L(\mathbf{x}_k, \boldsymbol{\lambda}_k)$ is needed (in many applications the Hessian is not easy to compute), this is done here using the *BFGS method* for quasi-Newton approximation. This procedure can be found in [21] Section 18.3. The new iterate is finally computed by the algorithm as $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$, where α_k is chosen in order to have a sufficient decrease in an appropriate chosen merit function, for more details about how this function is chosen see the function description in the Matlab Product Help.

Bibliography

[1] C. Acerbi, D. Tasche. *On the coherence of expected shortfall.* Journal of Banking and Finance 26:1487-1503, 2002.

[2] P. Artzner, F. Delbaen. *Coherent measures of risk.* Journal of Mathematical Finance 9:203-228, 1999.

[3] A.K. Bera, S.Y. Park. *Optimal portfolio diversification using maximum entropy.* Working Paper, 2004.

[4] M. Best, R. Grauer. *On the sensitivity of the mean-variance-efficient portfolios to changes in asset means: some analytical and computational results.* Review of Financial Studies 4:315-342, 1991.

[5] M. Denault. *Coherent allocation of risk capital.* Journal of Risk 4(1):7-21, 2001.

[6] R. Durrett. *Probability: Theory and Examples.* Wadsworth, Belmont, Second Edition, 1995.

[7] P. Embrechts, D. Lambrigger, M. Wüthrich. *Multivariate extremes and the aggregation of dependent risks: examples and counter-examples.* Extremes 12:107-127, 2008.

[8] P. Embrechts, J. Neslehova, M. Wüthrich. *Additivity properties for Value-at-Risk under Archimedean dependence and heavy-tailedness.* Insurance: Mathematics and Economics 44:164-169, 2009.

[9] L. Hannah, J.A. Kay. *Concentration in Modern Industry. Theory, Measurement and the UK Experience.* Macmillan, 1977.

[10] J.E. Ingersoll. *Theory of Financial Decision Making.* Rowman Littlefield, 1987.

[11] O. Ledoit, M. Wolf. *Improved estimation of the covariance matrix of stock returns with an application to portfolio selection.* Journal of Empirical Finance 10:603-621, 2003.

[12] O. Ledoit, M. Wolf. *Honey, I shrunk the sample covariance matrix.* Journal of Portfolio Management 30:110-119, 2003.

[13] S. Maillard, T. Roncalli, J. Teiletche. *On the properties of equally-weighted risk contributions portfolios.* Working Paper, 2009.

[14] H.M. Markowitz. *Portfolio selection.* Journal of Finance 7(1):77-91, 1952.

[15] H.M. Markowitz. *Portfolio Selection: Efficient Diversification of Investments.* Wiley, 1959.

[16] A.J. McNeil, R. Frey. *Estimation of tail-related risk measures for heterostatic financial time series: an extreme value approach.* Journal of Empirical Finance 7:271-300, 2000.

[17] A.J. McNeil, R. Frey, P. Embrechts. *Quantitative Risk Management: Concepts, Techniques, and Tools.* Princeton University Press, 2005.

[18] A. Meucci. *Risk and Asset Allocation.* Springer, 2005.

[19] A. Meucci. *Managing diversification.* Risk 22:74-79, 2009.

[20] Y. Muromachi. *Decomposing total risk of a portfolio into the contributions of individual assets.* Working Paper, 2009.

[21] J. Nocedal, S.J. Wright. *Numerical Optimization.* Springer, Second Edition, 2000.

[22] E. Qian. *Risk Parity Portfolios: Efficient Portfolios Through True Diversification.* White Paper, 2005.

[23] T. Rockafellar, S. Uryasev. *Conditional value-at-risk for general loss distributions.* Journal of Banking and Finance 26:1443-1471, 2002.

[24] D. Tasche. *Risk contributions and performance measurement.* Preprint, Department of Mathematics, TU-München, 1999.

[25] D. Tasche. *Conditional expectation as quantile derivative. Risk Contributions and Performance Measurement.* Preprint, Department of Mathematics, TU-München, 2000.